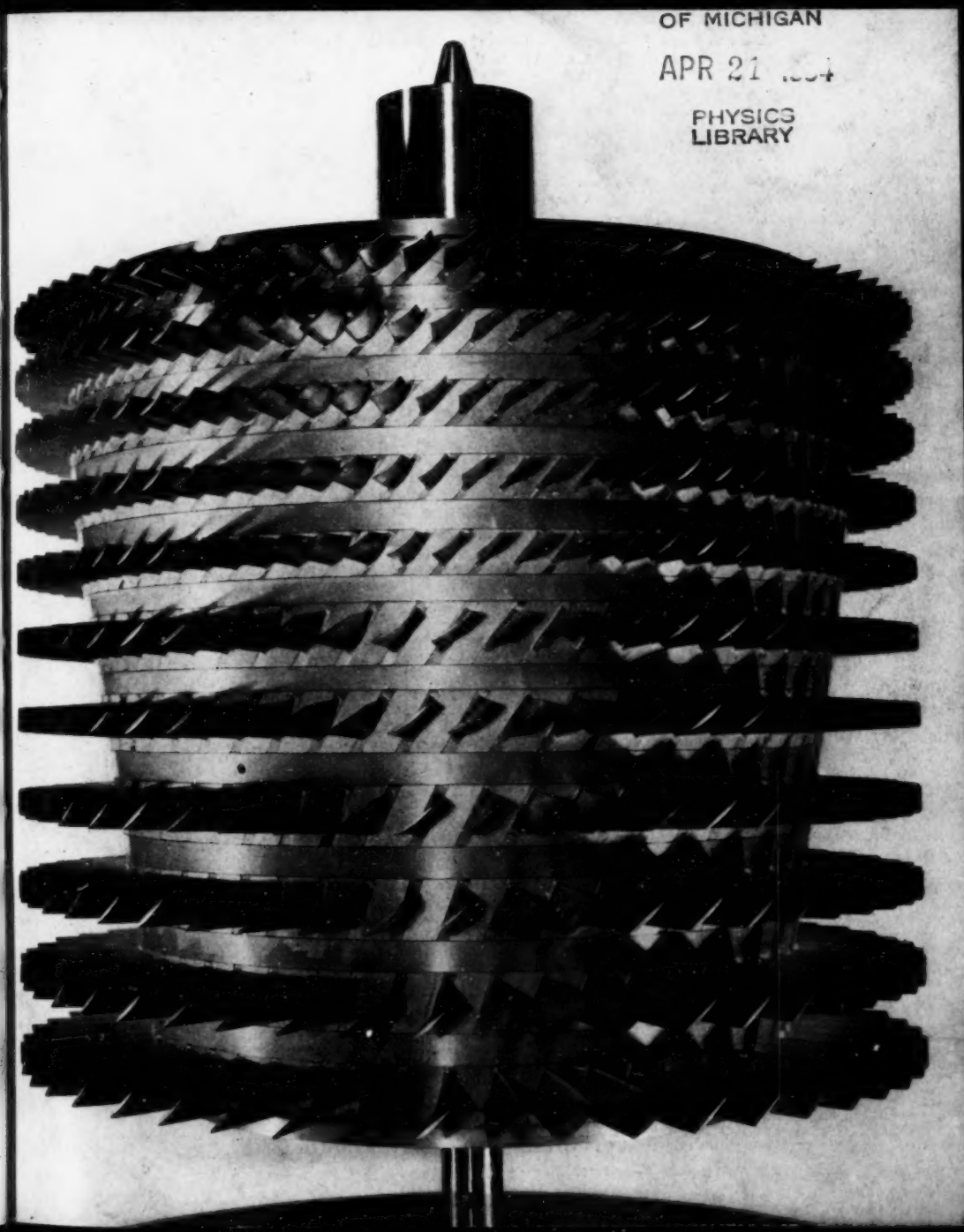
JOURNAL OF APPLIED PHYSICS

OF MICHIGAN
APR 21
PHYSICS
LIBRARY



JOURNAL OF APPLIED PHYSICS

Vol. 25, No. 4 APRIL, 1954

In This Issue

	Page
Electromagnetic Velometry. II. Elimination of the Effects of Induced Currents in Explorations of the Velocity Distribution in Axially Symmetrical Flow — Alexander Kolin and Fritz Reiche	409
Determination of the Resultant Dipole of the Heart from Measurements on the Body Surface — D. Gabor and C. V. Nelson	413
Secondary Electron Resonance Mechanism of Low-Pressure High-Frequency Gas Breakdown — Albert J. Hatch and H. Bartel Williams	417
Switching in Bistable Circuits — R. Stuart Mackay	424
Dynamics of Corona Discharge between Cylindrical Electrodes — L. Colli, U. Facchini, E. Gatti, and A. Persano	429
Focusing of an Electron Beam by Periodic Fields — A. M. Clogston and H. Heffner	436
Germanium Diodes from Spherical Pellets — W. C. Dunlap, Jr.	448
Studies on the Creep Recovery and Annealing of Zinc Single Crystals — M. Tanenbaum and W. Kauzmann	451
Resistivity Striations in Germanium Crystals — Paul R. Camp	459
Motion of Small Particles in Skew Shape Suspended in a Viscous Liquid — Chan-Mou Tchen	463
Growth of a Vapor Bubble in a Superheated Liquid — H. K. Forster and N. Zuber	474
Magnetic Switching Circuits — Robert C. Minnick	479
Graphical Method of Evaluating Trigonometric Functions Used in Crystal Structure Analysis. Parts I and II — H. J. Grenville-Wells	485
Necking of a Tension Specimen in Plane Plastic Flow — E. T. Onat and W. Prager	491
Growth of Vapor Bubbles in Superheated Liquids — M. S. Plesset and S. A. Zwick	493
Contact Charging Between a Borosilicate Glass and Nickel — John W. Peterson	501
Estimating Differential Scattering Cross Sections in Gases — Tino Ahrens	505
Backscattering of the Co^{60} Gamma Rays from Infinite Media — Evans Hayward and John H. Hubbell	506
Interpretation of Photoelastic Transmission Patterns for a Three-Dimensional Model — D. C. Drucker and W. B. Woodward	510
Point Emitter-Junction Collector Transistor — Robert H. Kingston	513
Concept of Stability for Nuclear Reactors — Lawrence Baylor Robinson	516
Generalized Off-Axis Distributions from Disk Sources of Radiation — Jack H. Smith and Martin L. Storm	519
Propagation of Longitudinal Deformation Waves in a Prestressed Rod of Material Exhibiting a Strain-Rate Effect — Robert J. Rubin	528
Measurements of the Gas Temperature of a Low-Pressure RF Discharge — Irving Reingold and Kenton Garoff	537
Letters to the Editor:	
On the Diffraction of X-Rays by a Twisted Crystal Lattice — S. S. Hsu	541
Electromagnetic Field near a Dielectric Lens — D. C. Hogg	542
Application of Perturbation Theory to the Traveling Wave Tubes — M. W. Muller and W. L. Beaver	542
Regrowth of Silicon through a Low Melting Zone of Silicon-Gold Eutectic — J. N. Carman, P. E. Stello, and C. A. Bittmann	543
Erratum: Approximate Solutions of the Space-Charge Problem for Some Unusual Electrode Geometries — Henry F. Ivey	543
Errata: The Vector Wave Function Solution of the Diffraction of Electromagnetic Waves by Circular Disks and Apertures. II. The Diffraction Problems — Carson Flammer	543
Errata: Sudden Change in the Arrangement of Crystallites in Thin Evaporated Bismuth Films — Lorn L. Howard	543
Announcement	544
Index to Advertisers	xxiv

ELMER HUTCHISSON, *Editor*

EARLE C. GREGG, *Assistant Editor*

Board of Editors

MAURICE A. BIOT
RAY E. BOLZ
WALLACE J. ECKERT
ROBERT D. HEIDENREICH
LOUIS A. PIPES
ROMAN SMOLUCHOWSKI
W. R. SMYTHE
ARTHUR V. TOBOLSKY
ALVIN M. WEINBERG

Publication of the American Institute of Physics

OFFICERS

FREDERICK SEITZ, *Chairman*
GEORGE B. PEGRAM, *Treasurer*

GOVERNING BOARD

ROBERT F. BACHER
KARL K. DARROW
S. A. GOUDSMIT
DEANE B. JUDD
HUGH S. KNOWLES
C. C. LAURITSEN
W. H. MARKWOOD, JR.
WILLIAM F. MEGGERS
PHILIP M. MORSE
BRIAN O'BRIEN
HARRY F. OLSON
GEORGE B. PEGRAM
ERIC RODGERS
RALPH A. SAWYER
FREDERICK SEITZ
WILLIAM SHOCKLEY
RICHARD M. SUTTON
J. H. VAN VLECK
MARK W. ZEMANSKY

Consultant on Printing
MELVIN LOOS

ADMINISTRATIVE STAFF

HENRY A. BARTON, *Director*
WALLACE WATERFALL, *Executive Secretary*
THEODORE VORBURGER, *Advertising Manager*
RUTH F. BRYANS, *Publication Manager*
EDITH I. NEFTEL *Circulation Manager*
KATHRYN SETZE *Assistant Treasurer*

The JOURNAL OF APPLIED PHYSICS, published monthly at Prince and Lemon Sts., Lancaster, Pa., is devoted to physics in its role as the science basic to other natural sciences and to the arts and industries. Previous to 1937 this publication was known as PHYSICS. The JOURNAL OF APPLIED PHYSICS publishes editorials and reviews, as well as technical papers of applied physics. Articles appearing in it are indexed regularly, according to their field, by Physics Abstracts, Chemical Abstracts, Engineering Index, Inc., Industrial Arts Index, etc.

Manuscripts should be submitted to Elmer Hutchisson, Editor, Case Institute of Technology, Cleveland 6, Ohio. Unless otherwise stated, submission of a manuscript is a representation that it has been neither copyrighted, published, nor submitted for publication elsewhere. The authors' institutions are requested to pay a **publication charge** of $8 per page which, if honored, entitles them to 100 reprints without further charge. Instructions will be sent with galley proofs.

Subscriptions, renewals, and orders for back numbers should be addressed to the American Institute of Physics, 57 East 55 Street, New York 22, New York.

Subscription Price	U.S. and Canada	Elsewhere
To members of the American Institute of Physics and Affiliated Societies	$10.00	$11.00
To all others	12.00	13.00

Back Number Prices

Yearly back number rate when complete year is available: $14.00. Single copies: $1.25 each.

Changes of address should be sent to the American Institute of Physics; **corrected proofs** should be sent to the Publication Manager, American Institute of Physics. **Advertising rates** supplied on request. Orders, advertising copy, and cuts should be sent to the American Institute of Physics.

Entered as second class matter January 22, 1937, at the Post Office at Lancaster, Pennsylvania, under the Act of March 3, 1879. Accepted for mailing at the special rate of postage provided for in paragraph (d-2), section 34.40, P. L. & R. of 1948, authorized May 2, 1932.

Journal
of
Applied Physics

Volume 25, Number 4 April, 1954

Electromagnetic Velometry. II. Elimination of the Effects of Induced Currents in Explorations of the Velocity Distribution in Axially Symmetrical Flow*

ALEXANDER KOLIN, *University of Chicago, Chicago, Illinois*

AND

FRITZ REICHE, *New York University, University Heights, New York*

(Received May 28, 1953)

The distribution of electrical potential in a conduit carrying an axially symmetrical flow through an homogeneous magnetic field, which is perpendicular to the flow, is analyzed. It is shown that it is possible to eliminate the difficulties arising from induced currents in electromagnetic determinations of the velocity distribution by measuring the potential gradient $\partial\varphi/\partial y$ (y being $\perp H$ and $\perp v$) along a diameter bisecting the angle between the H axis and the Y axis. The knowledge of the distribution of $(\partial\varphi/\partial y)_{45°}$ at a known discharge through the conduit determines the velocity distribution $v=f(r)$.

INTRODUCTION

IN an earlier paper one of the authors[1] described a method using electomagnetic induction for the determination of the velocity distribution in a moving fluid. The fluid is allowed to move across an homogeneous magnetic field; as a result, an emf is induced in the moving fluid. If the velocity is not uniform, it is possible to establish the velocity distribution function by measuring the induced potential gradient at desired points. In the case of a nonuniform velocity distribution, however, there is an essential complication. Electrical eddy currents are usually set up wherever there is a velocity gradient. The potential difference picked up by the two prongs of a "probe" (1) is therefore determined not only by the local instantaneous velocity of flow, but is also influenced by the current density at the point in question, which will, in general, depend upon the flow conditions in the environment. For the performance of local velocity measurements, two lines of approach remain open: (1) One can eliminate the currents at the point of measurement by trapping the fluid volume momentarily in a small insulating chamber which shields it from extraneous currents while the potential difference induced across it is being measured ("shielded electrode").[1] (2) The alternative procedure would consist of refraining from shielding the electrode tips and in finding a theoretical way of correcting for the flow of induced currents.

Although the first method has been worked out satisfactorily, the development of the second method is very desirable since it is hydrodynamically superior to the first method. The omission of the shield permits the use of very fine wires as electrodes which present a smaller obstacle to flow than a shielded sensing unit. Calculations carried out below for a few practically important cases have led to the suggestion of a simple method of correcting for the flow of induced currents by a special procedure of measurement, rather than by computation.

The following considerations will be limited to uniform, axially symmetrical flow through straight circular pipes and to circular jets emerging into a circular conduit.

FORMULATION OF THE PROBLEM

We shall consider the case of a uniform circular nonconducting conduit through which a conducting fluid flows in the direction of the positive Z axis. The velocity distribution is symmetrical with respect to the Z axis and independent of the angle θ which is one of the two polar co-ordinates in the $X-Y$ plane. The pipe is in an infinitely extended homogeneous magnetic field which is directed parallel to the X axis. According to Maxwell's

* This publication has been aided by the Abbott Memorial Fund of the University of Chicago.

[1] A. Kolin, J. Appl. Phys. 15, 150–164 (1944).

1st equation for a moving medium,

$$\text{curl}\mathbf{E}=-\frac{\mu}{c}\frac{\partial\mathbf{H}}{\partial t}+\frac{\mu}{c}\text{curl}[\mathbf{v}\times\mathbf{H}]. \tag{1}$$

Since in our method we measure the induced emf either with a constant magnetic field or with an alternating magnetic field at a moment when $H=H_{max}$: $\partial H/\partial t=0$.

Hence,

$$\text{curl}\left\{\frac{\mu}{c}[\mathbf{v}\times\mathbf{H}]-\mathbf{E}\right\}=0. \tag{2}$$

The content of the parenthesis can be expressed as the gradient of a potential function

$$\frac{\mu}{c}[\mathbf{v}\times\mathbf{H}]-\mathbf{E}=\text{grad}\varphi, \tag{3}$$

where φ is the potential with respect to the center of the pipe.

In our special case

$$\mu=1;\quad v_x=f(r);\quad v_z=v_y'=0;$$
$$H_z=H;\quad H_y=H_x=0;\quad r=(x^2+y^2)^{\frac{1}{2}};\quad \partial\varphi/\partial z=0.$$

Under these circumstances the component equations for (3) become:

$$\text{(a)}\quad E_x=-\partial\varphi/\partial x,$$
$$\text{(b)}\quad E_y=-\partial\varphi/\partial y+\frac{H}{c}f(r), \tag{4}$$
$$\text{(c)}\quad E_z=0.$$

Equation (4b) is the most important one for us since it contains the local velocity $f(r)$.

By substituting the values of Eq. (4) in

$$\text{div}\mathbf{E}=0 \tag{5}$$

we get

$$\partial E_x/\partial x+\partial E_y/\partial y+\partial E_z/\partial z=0$$
$$=-\frac{\partial^2\varphi}{\partial x^2}-\frac{\partial^2\varphi}{\partial y^2}-\frac{H}{c}\frac{\partial}{\partial r}f(r)\frac{\partial r}{\partial y}, \tag{5a}$$

or

$$\frac{\partial^2\varphi}{\partial x^2}+\frac{\partial^2\varphi}{\partial y^2}=\frac{H}{c}f'(r)\frac{y}{r}. \tag{5b}$$

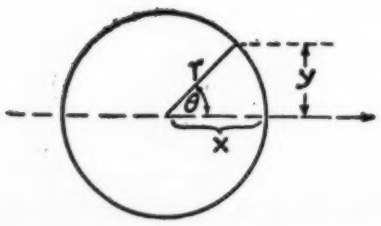


FIG. 1.

In polar coordinates

$$\frac{\partial^2\varphi}{\partial r^2}+\frac{1}{r}\frac{\partial\varphi}{\partial r}+\frac{1}{r^2}\frac{\partial^2\varphi}{\partial\theta^2}=\frac{H}{c}f'(r)\sin\theta. \tag{6}$$

This is the differential equation of our problem for axially symmetrical steady flow across a perpendicular magnetic field. (See Fig. 1.) In the following paragraph this equation will be integrated under appropriate boundary conditions. The analysis of the solution will suggest a method of measurement in which the induced currents do not enter the picture.

INTEGRATION OF EQ. (6)

We assume a solution of the form $\varphi=\psi(r)F(\theta)$ and more specifically

$$\varphi=\psi(r)\sin\theta. \tag{8}$$

The boundary conditions are as follows: (A) The velocity vanishes at the wall: $f(R)=0$. (B) The radial components of current and electric field intensity vanish at the wall: $i_R=0=E_R$. (C) φ is finite everywhere.

Substituting (8) in (6),

$$\frac{d^2\psi}{dr^2}+\frac{1}{r}\frac{d\psi}{dr}-\frac{1}{r^2}\psi=\frac{H}{c}f'(r);$$

First integral:

$$\frac{d\psi}{dr}+\frac{\psi}{r}=\frac{H}{c}f(r)+2C_1; \tag{9}$$

Second integral:

$$\psi=\frac{H}{c}\frac{1}{r}\int_0^r rf(r)dr+C_1r+\frac{C_2}{r}. \tag{10}$$

$C_2=0$ because of boundary condition (C). C_1 follows from boundary conditions (A) and (B); since $\partial\psi/\partial r=0$ for $r=R$ and $f(R)=0$, we get from (10):

$$C_1=\frac{H}{c}\frac{1}{R^2}\int_0^R rf(r)dr. \tag{11}$$

Hence,

$$\varphi=\frac{H}{c}\sin\theta\left\{\frac{1}{r}\int_0^r rf(r)dr+\frac{r}{R^2}\int_0^R rf(r)dr\right\}. \tag{12}$$

INTERPRETATION OF THE INTEGRALS

$$\frac{1}{R^2}\int_0^R rf(r)dr=\frac{1}{2}\frac{\int_0^R 2\pi rf(r)dr}{\pi R^2}$$
$$=\tfrac{1}{2}\bar{V}_R:\tfrac{1}{2}\text{ average speed of flow through the pipe.} \tag{13}$$

$$\frac{1}{r^2}\int_0^r rf(r)dr=\frac{1}{2}\frac{\int_0^r 2\pi rf(r)dr}{\pi r^2}$$
$$=\tfrac{1}{2}\bar{V}_r:\tfrac{1}{2}\text{ average speed of flow through imaginary cylinder of radius } r.$$

Thus we can write

$$\varphi=\frac{H}{c}y\frac{\bar{V}_r+\bar{V}_R}{2}=\frac{H}{c}y\bar{V}_r, \tag{12a}$$

where

$$\bar{V}_r=\frac{\bar{V}_r+\bar{V}_R}{2}$$

is the average of the average speeds $\bar{V}_r$ and $\bar{V}_R$.

RELATION BETWEEN THE POTENTIAL GRADIENT AND LOCAL VELOCITY

From (12a), (13), and (14),

$$\frac{\partial\varphi}{\partial y}=\frac{H}{2c}(f(r)+\bar{V}_R+\cos 2\theta[\bar{V}_r-f(r)]). \tag{15}$$

The angle θ determines the line along which the exploring electrodes are moved while the potential gradient $\partial\varphi/\partial y$ is being measured. When we proceed along the Y axis: $\theta=90°$ and then

$$\left(\frac{\partial\varphi}{\partial y}\right)_{90°}=\frac{H}{c}\left[f(r)-\frac{\bar{V}_r-\bar{V}_R}{2}\right],$$

or

$$f(r)=\frac{c}{H}\left(\frac{\partial\varphi}{\partial y}\right)_{90°}+\left(\frac{\bar{V}_r-\bar{V}_R}{2}\right). \tag{16}$$

This equation for the local velocity $f(r)$ has a disadvantage: in addition to measuring $\partial\varphi/\partial y$ at every point, we have to determine a variable correction term $(\bar{V}_r-\bar{V}_R/2)$. This term vanishes at the wall for $r=R$ and is usually positive since normally $\bar{V}_r>\bar{V}_R$. It is especially small for a turbulent velocity distribution. In the limiting case of a uniform velocity distribution, $\bar{V}_r=\bar{V}_R$ for all values of r. Thus, the correction becomes negligible for highly turbulent flows approaching this condition.

The procedure of establishing the velocity distribution is greatly simplified if we move the probe along a line of $\theta=45°$ while measuring $\partial\varphi/\partial y$. Then,

$$\left(\frac{\partial\varphi}{\partial y}\right)_{45°}=\frac{H}{2c}(f(r)+\bar{V}_R),$$

or

$$f(r)=\frac{2c}{H}\left(\frac{\partial\varphi}{\partial y}\right)_{45°}-\bar{V}_R. \tag{17}$$

Here the correction term is constant (being equal to the average speed of flow) and can be determined quite simply either by a hydraulic determination of the average flow from the discharge rate or electrically by measuring the potential difference across the pipe diameter parallel to the Y axis:

$$(\varphi_R)_{\pi/2}-(\varphi_R)_{3\pi/2}=\frac{H}{c}2R\bar{V}_R.$$

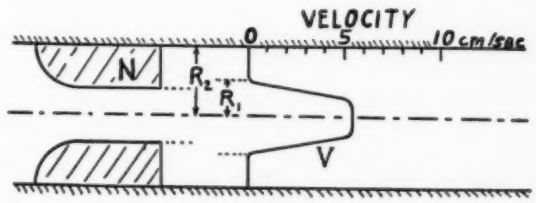


FIG. 2. Velocity distribution V in the pipe some distance behind the nozzle N.

This potential difference is proportional to the average velocity for any axially symmetrical velocity distribution, as has been shown previously.[2,3]

This simple procedure of measuring the velocity distribution by proceding along a diameter forming an angle of 45° with the magnetic field is valid not only for flow filling the entire cross section of the pipe but is equally applicable to a jet escaping into a pipe since the only condition we imposed upon the velocity distribution was central symmetry and the jet could be considered as a special case of a symmetrical velocity distribution (Fig. 2). In this case, however, the expression (17) has to be interpreted as follows: in the interval of the jet $(0<r<R_1)$, the gradient of the potential φ_1: $(\partial\varphi_1/\partial y)_{45°}=H/2c\cdot(f(r)+\bar{V}_{R_2})$ varies as a function of r. In the interval outside the jet $(R_1<r\leqq R_2)$, the gradient of the potential φ_2:

$$(\partial\varphi_2/\partial y)_{45°}=\frac{H}{2c}\bar{V}_{R_2}=\text{const},$$

where

$$\bar{V}_{R_2}=\frac{1}{\pi R_2^2}\int_0^{R_1}2\pi r f(r)dr=\frac{Q}{\pi R_2^2},$$

Q being the liquid volume transported by the jet per second (Fig. 3a).

$\bar{V}_{R_2}$ can be determined hydraulically or electrically in the same manner as $\bar{V}_R$ of Eq. (17).† There is, however, another electrical way of determining $\bar{V}_R$ for a pipe as well as $\bar{V}_{R_2}$ for a jet which may prove to be of particular practical importance. From Eq. (17) we see that at the wall where $f(r)=0$,

$$\bar{V}_R=\frac{2c}{H}(\partial\varphi/\partial y)_{45°}. \tag{18}$$

---

[2] B. Thürlemann, Helv. Phys. Acta 14, 383–419 (1941).
[3] A. Kolin, Rev. Sci. Instr. 16, 109–116 (1945).

† If $\bar{V}_{R_2}$ shall be determined electrically, one needs the expressions for the potentials in the two regions depicted in Fig. 3(a):

Region 1: $0\leqq r<R_1$
Region 2: $R_1<r\leqq R_2$

$$\left.\begin{aligned}\varphi_1&=\frac{H}{2c}\sin\theta\left(\bar{V}_r+\bar{V}_{R_2}\right)r\\ \varphi_2&=\frac{H}{2c}\sin\theta\left(r+\frac{R_2^2}{r}\right)\bar{V}_{R_2}\end{aligned}\right\} \tag{17a}$$

Hence:

$$(\varphi_2)_{\substack{r=R_2\\ \theta=\pi/2}}-(\varphi_2)_{\substack{r=R_2\\ \theta=3\pi/2}}=\frac{2H}{c}R_2\bar{V}_{R_2}.$$

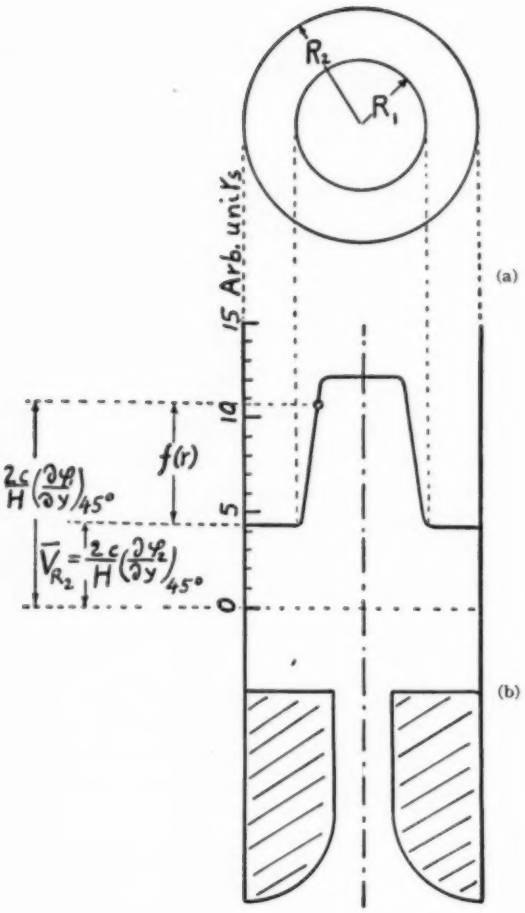


FIG. 3(a). The jet region (R_1) and the annular region of zero velocity (limited by R_1 and R_2). (b). Determination of the local velocity $f(r)$ from the measured quantities (see Fig. 2).

Two important conclusions follow from this fact: (1) The additive constant $\bar{V}_R$ of Eq. (17) is obtained from the reading $(H/2c)\bar{V}_R$ at the wall ($r=R$) when we measure $(\partial\varphi/\partial y)_{45°}$ between 0 and R. This reading is obviously not zero, but, rather, depends on the average velocity. (2) The average velocity in a pipe can be measured by introducing two closely spaced electrodes through an opening in the pipe wall at the end of a diameter which forms an angle of 45° with the magnetic field lines. The tips of these electrodes lie along a vertical line so as to measure $\partial\varphi/\partial y$. This method may, in some cases, be preferable to the original method of measuring the potential difference between the ends of a diameter perpendicular to the magnetic field lines, which requires perforation of the pipe in two widely separated places.

In the case of the jet problem, we can see from Eq. (17), taking into account the fact that $f(r)=0$ and $\varphi=\varphi_2$ for $R_1<r<R_2$, $\varphi=\varphi_1$ for $0<r<R_1$ and $\bar{V}_R=\bar{V}_{R_2}$:

$$\bar{V}_{R_2}=\frac{2c}{H}(\partial\varphi_1/\partial y)_{R_1;45°}=\frac{2c}{H}(\partial\varphi_2/\partial y)_{45°}. \qquad (19)$$

Hence, in measuring $(\partial\varphi/\partial y)_{45°}$ the instrument deflection will reach a constant value at $r=R_1$ which will persist until $r=R_2$. The ordinate of this horizontal part of the $(\partial\varphi/\partial y)=F(r)$ function determines the additive constant $\bar{V}_{R2}$ and hence also the discharges through the nozzle.

Figure 3(b) illustrates how the velocity distribution $f(r)$ in the jet can be inferred, according to Eq. (17), from graph representing the measured $(\partial\varphi/\partial y)_{45°}$ as a function of the distance from the axis of symmetry.

It is interesting to note that $(\partial\varphi_2/\partial y)_{R_2;45°}=-(\partial\varphi_2/\partial x)_{R_2;45°}$.‡ This holds for a pipe as well as for the jet problem or any axially symmetrical velocity distribution.

THE POTENTIAL GRADIENT IN A PIPE AS A FUNCTION OF RADIAL POSITION AS DETERMINED ALONG DIAMETERS OF DIFFERENT INCLINATION

The function $(\partial\varphi/\partial y)_{\theta;r}=F(r,\theta)$ for $\theta=$const has a different appearance according to the choice of θ; (a corresponding statement can be made about $\partial\varphi/\partial x$). As an example, we shall examine the cases of $\theta=90°$; $\theta=0°$, and $\theta=45°$.

(A) $\theta=90°$: From Eq. (12) and (15) follows

$$\partial\varphi/\partial x=0 \quad \text{and} \quad \partial\varphi/\partial y=\frac{H}{c}\left[f(r)+\frac{\bar{V}_R-\bar{V}_r}{2}\right],$$

(B) $\theta=0°$:

$$\partial\varphi/\partial x=0 \quad \text{and} \quad \partial\varphi/\partial y=\frac{H}{c}\left[\frac{\bar{V}_R+\bar{V}_r}{2}\right],$$

(C) $\theta=45°$:

$$\partial\varphi/\partial x=-\frac{H}{2c}[\bar{V}_r-f(r)]; \quad \partial\varphi/\partial y=\frac{H}{2c}[f(r)+\bar{V}_R].$$

Of all these expressions the only practical equation for the determination of $f(r)$ is the one for $\partial\varphi/\partial y$ at $\theta=45°$ since it contains in addition to $\partial\varphi/\partial y$ and $f(r)$ only a constant correction which can be easily determined.

It is interesting to note that, when getting $(\partial\varphi/\partial y)$, we obtain two different curves depending on whether we move our prongs along the X axis or along the Y axis (cases (A) and (B)).

‡ From (17a) follows:

$$\varphi_2=\frac{H}{2c}\bar{V}_{R_2}y\left[1+\left(\frac{R_2}{r}\right)^2\right].$$

Hence,

$$\partial\varphi_2/\partial x=-\frac{H}{2c}\bar{V}_{R_2}\frac{2R_2^2xy}{r^4}$$

and

$$\partial\varphi_2/\partial y=\frac{H}{2c}\bar{V}_{R_2}\left[1+\left(\frac{R_2}{r}\right)^2-2\frac{R_2^2y^2}{r^4}\right].$$

Thus, $(\partial\varphi_2/\partial x)_{R_2;45°}=-\frac{H\bar{V}_{R_2}}{2c}$ and $(\partial\varphi_2/\partial y)_{R_2;45°}=+\frac{H\bar{V}_{R_2}}{2c}$.

The best method is obviously to measure along the axis which bisects the angle between the X axis and the Y axis (case C); comp. Eq. (17). It is, however, also possible to obtain $f(r)$ from the curves corresponding to the above equations (A) and (B). From (A) and (B) it follows that

$$(\partial\varphi/\partial y)_{r;90°}+(\partial\varphi/\partial y)_{r;0°}=\frac{H}{c}[f(r)+\bar{V}_R].$$

Hence,

$$f(r)=\frac{c}{H}[(\partial\varphi/\partial y)_{r;0°}+(\partial\varphi/\partial y)_{r;90°}]-\bar{V}_R. \tag{20}$$

Equation (17) and, as a less elegant alternative, Eq. (20) enable us to determine the velocity distribution in axially symmetrical flow by electrical measurements. The measurement involves the determination of $(\partial\varphi/\partial y)$ at points located on a predetermined diameter ($\theta=45°$ for Eq. (17) and $\theta_1=0°$; $\theta_2=90°$ for Eq. (20)) and one measurement of $\bar{V}_R$. $\bar{V}_R$ follows from the expression for the potential difference across the pipe diameter parallel to the Y axis: $\Delta\varphi=(H/c)2R\bar{V}_R$.

Hence,

$$\bar{V}_R=\frac{c}{H}\frac{\Delta\varphi}{2R}. \tag{21}$$

The determination of $(\partial\varphi/\partial y)$ amounts practically to measurement of the local potential difference existing between two points located along a line parallel to the Y axis and separated by a distance negligibly small as compared to the pipe diameter. Electrodes suitable for such measurements have been described in reference 1. Because of their extreme fineness, this method offers the possibility of studying the velocity distribution with a negligible disturbance of the flow.

JOURNAL OF APPLIED PHYSICS VOLUME 25, NUMBER 4 APRIL, 1954

Determination of the Resultant Dipole of the Heart from Measurements on the Body Surface

D. GABOR AND C. V. NELSON*†
Department of Electrical Engineering, Imperial College of Science and Technology, London, England, and Physiology Department, Middlesex Hospital Medical School, London, England

(Received August 14, 1953)

By means of the vector calculus, it is proved that the magnitude, orientation, and location of the resultant dipole of a system of sources and sinks inside a finite volume conductor is given by an integration over the bounding surface. The method is applied to finding the "heart vector," or the resultant dipole moment of the human heart. The theory was checked in two- and three-dimensional electrolytic tank models of the human thorax.

MANY problems of electrophysiology are concerned with the distribution of current and potential due to sources and sinks of electricity inside a volume conductor. Thus, in electrocardiology, one measures the potentials on the surface of the body produced by the electrical action of the heart. The theoretical analysis involves the determination of the "heart vector," which is defined as the resultant dipole of all the individual sources and sinks in the heart. These analyses all assume that the thorax can be replaced by some simple geometric figure. The purpose of the present paper is to explain the mathematical and physical basis of a new method which makes it possible to determine the resulting dipole of the heart *and its location* from measurements on the surface, without any special assumptions as to the shape of the body.

It is well known that it is not possible to determine in detail the sources and sinks of currents even in a body of exactly known constitution (that is to say a body whose conductivity is given as a function of the position) from potential measurements carried out on the surface. If we can account for the observed surface potentials by one set of sources and sinks, we can construct an infinite number of other distributions which account for it equally well, by adding any number of sources and equal sinks so designed that there is no current outside a certain closed surface. Yet it will be shown that the resultant dipole of the heart, its orientation and location, can be determined from measurements carried out on the body surface in a satisfactory way. Complications arise only if one wants to determine the system of sources and sinks in more detail, for instance if one wants to replace the heart by two or more dipoles.

The mathematical theory will be given only for a body of uniform isotropic conductivity k. It was found, however, that the simple theory gives satisfactory accuracy also in models of the human body in which the heart is almost surrounded by the lungs, whose conductivity is only about a quarter of that of the rest.

* Research Fellow of the American Heart Association.
† Now at the Heart Station, Department of Internal Medicine, University of Utah, Salt Lake City, Utah.

Let $\mathbf{i}$ be the vector of current density, and $s=\nabla\cdot\mathbf{i}$ the source strength, whose sum extended over the whole body volume is nil. The resultant vector dipole moment of such a source system is defined as

$$\mathbf{M}=\int\int\int \mathbf{r}s dv, \tag{1}$$

where $\mathbf{r}$ is the radius vector from an arbitrary origin, and dv the volume element. Consider now e.g., the x-component of this vector. We transform this as follows:

$$\begin{aligned} M_x&=\int\int\int xs dv=\int\int\int x\nabla\cdot\mathbf{i}dv \\ &=\int\int\int\nabla\cdot(x\mathbf{i})dv-\int\int\int\mathbf{i}\cdot\nabla x dv \\ &=\int\int xi_n dS-\int\int\int i_x dv=k\int\int\int\frac{\partial V}{\partial x}dv \\ &=k\int\int V dy dz=k\int\int V dS_x. \end{aligned} \tag{2}$$

In the first line we have used the theorem on the divergence of the product of a scalar and a vector. In the first expression in the second line we have applied Gauss's theorem, and substituted the value of $\mathbf{i}\cdot\nabla x$. In the second expression we have taken account of the fact that the body surface is insulated, so that there is no normal component i_n of the current, and we have made use of Ohm's Law, $i_x=-k\partial V/\partial x$. In the next expression we have transformed this into a surface integral, by integrating over x. In the last expression we have put this into a convenient form, by introducing the vector surface element $d\mathbf{S}$, whose absolute value is dS and which points in the direction of the outer normal. We can sum up Eq. (2) together with two other component equations in the vector equation

$$\mathbf{M}=k\int\int V d\mathbf{S}. \tag{3}$$

Thus the strength and direction of the resultant dipole of any balanced current system in a uniformly conducting, insulated body can be determined by potential measurements on the surface.

Consider now the location of the resultant dipole. Note that if we have a point source and an equal sink, $+S$, $-S$, at $X+\frac{1}{2}\delta x$, $X-\frac{1}{2}\delta x$, $Y+\frac{1}{2}\delta y$, etc., the second moment of the source distribution is, in the x direction

$$\int\int\int x^2 s dv=S[(X+\tfrac{1}{2}\delta x)^2-(X-\tfrac{1}{2}\delta x)^2]=2X(S\delta x)$$

and passing to the limit, $\delta x\to 0$, $S\delta x\to M_x$, we have

$$XM_x=\frac{1}{2}\int\int\int x^2 s dv. \tag{4}$$

We now transform the second moment integral by similar steps as before

$$\begin{aligned} XM_x&=\frac{1}{2}\int\int\int x^2 s dv=\frac{1}{2}\int\int\int x^2\nabla\cdot\mathbf{i}dv \\ &=\frac{1}{2}\int\int\int\nabla\cdot(x^2\mathbf{i})dv-\frac{1}{2}\int\int\int\mathbf{i}\cdot\nabla x^2 dv \\ &=\frac{1}{2}\int\int x^2 i_n dS-\int\int\int x i_x dv \\ &=k\int\int\int x\frac{\partial V}{\partial x}dv=k\int\int xV dS_x \\ &\qquad -k\int\int\int V dv \end{aligned} \tag{5}$$

and we obtain two similar equations for YM_y and ZM_z. It is seen that we cannot determine any of these by the surface measurements, because we have the second term, with the volume integral of the potential V, in all three equations. However, in the same way as we have proved Eq. (4) we can derive

$$XM_y+YM_x=\int\int\int xys dv \tag{6}$$

and two similar equations, arising by cyclic permutation. One finds also, in just the same way as we have deduced Eqs. (2) and (5),

$$XM_y+YM_x=k\int\int V(x dS_y+y dS_x). \tag{7}$$

We can now eliminate the unknown volume integral of V between the three equations of the type (5), and together with the three equations of the type (7) we are left with five equations for the three quantities X, Y, Z. Any three of these five equations can be used for finding the location of the dipole, and the other two for checking how well the assumption of *one* dipole accounts for the observation. One can also use the method of least squares to obtain the best fit.

In the two-dimensional case, i.e., if one is interested only in the horizontal components and in the horizontal location of the heart dipole, there are no redundant equations. The moments of the first order give two equations;

$$M_x=k\int V dy,\quad M_y=k\int V dx, \tag{8}$$

and the moments of the second order give again two independent equations

$$XM_x - YM_y = k\int V(xdy - ydx),$$
$$XM_y + YM_x = k\int V(xdx + ydy). \quad (9)$$

These are just sufficient for determining the four data M_x, M_y, X, Y of the resultant dipole, and in the model experiments the results agreed very well with the actual data.

If one wants higher accuracy, and a more elaborate model of the heart, containing two, three, or more dipoles, the process can be continued. The moments of the third order of the source distribution give three equations of the type

$$\sum X^2 M_x = \frac{1}{3}\int\int\int x^3 s dv$$
$$= k\int\int x^2 V dS_x - 2k\int\int\int xV dv. \quad (10)$$

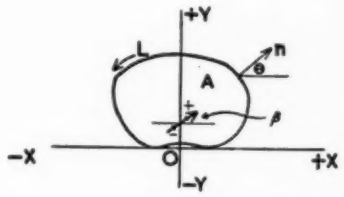


FIG. 1. Determination of dipole moment in two dimensions.

These contain the unknown first moments of the potential distribution. One can eliminate these by using the six further equations of the type

$$\sum (X^2 M_y + 2XYM_x)$$
$$= \frac{1}{2}\int\int\int x^2 y s dv$$
$$= k\int\int xyV dS_x - \int\int\int yV dv. \quad (11)$$

From Eq. (11) a second is derived by interchanging x and y, and from each of these two one obtains two other equations by cyclic interchange, altogether six equations of the type (11). These contain the same unknown first moments of V as the three Eqs. (10), so that one is left with six independent equations. To these is added the seventh and last

$$\sum (XYM_z + YZM_x + ZXM_y) = \int\int\int xyz s dv$$
$$= k\int\int V(xydS_z + yzdS_x + zxdS_y). \quad (12)$$

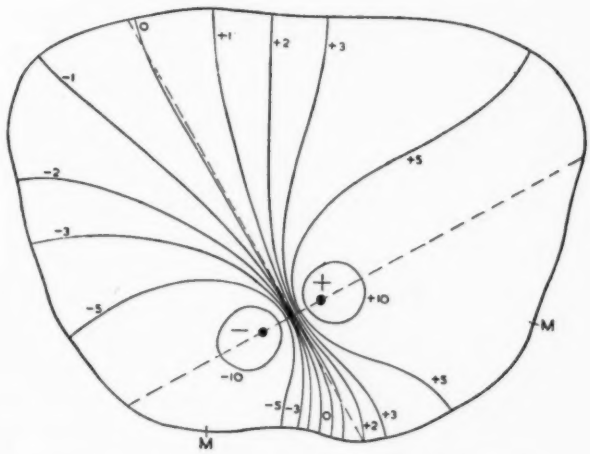


FIG. 2. Equipotential lines in a model of the thorax with uniform conductivity, $\beta = 30°$.

The summations at the left are to be carried out over the dipoles which one assumes in the model. It is seen that for one dipole one must go to second-order moments, giving two redundant equations, for two dipoles to the third order, giving $3+5+7-12=3$ redundancies. If one goes to moments of the fourth order, there are $3+5+7+9=24$ equations, so that the data are just sufficient for a model containing 4 dipoles, but the work of measurement and of calculation becomes prohibitive. This complication does not appear warranted before the clinical consequences of the "one dipole model" are exploited to the full.

APPLICATION OF THE METHOD TO FINDING THE RESULTANT HEART VECTOR

The theory was checked in electrolytic tank models made to represent the human thorax. In two dimensions, the model was based on a full-size cross section through the human thorax, and account was taken of

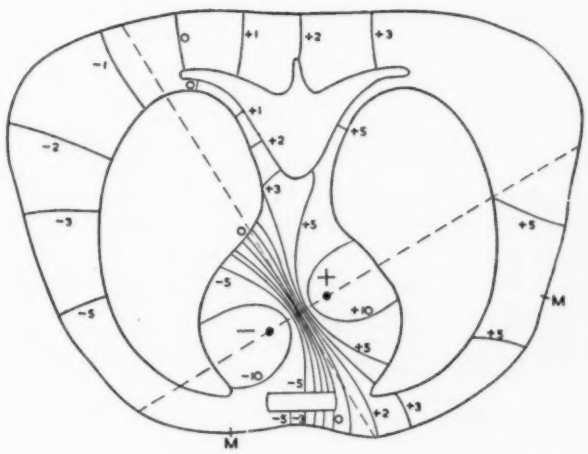


FIG. 3. Equipotential lines in a model of the thorax with insulating sternum and spine. The conductivity of the lungs is made $\frac{1}{4}$ of that of the rest. $\beta = 30°$.

TABLE I. Horizontal components of dipole moment for normal male subject.*

1 Time	2 M_x	3 M_y	4 M_x ma-cm	5 M_y ma-cm	6 M_h ma-cm	7 β degrees
R—start	$-47k$	$-239k$	−0.094	−0.478	0.49	−101
R—¼	$20k$	$-44k$	0.040	−0.088	0.095	−65
R—peak	$108k$	$218k$	0.216	0.437	0.49	64
R—¾	$-145k$	$370k$	−0.290	0.740	0.79	111
R—end	$-200k$	$435k$	−0.400	0.870	0.96	115
S—end	$-40k$	$75k$	−0.080	0.150	0.17	118

* Column 1 is the time during the RS complex of lead II, columns 2 and 3 give the values of M_x and M_y in terms of the conductivity k. If it is assumed that the average resistivity of the thorax interior is about 500 ohm-cm, then $k=1/500$ mho-cm^{-1}. Using this value of k, and with the potential V expressed in millivolts, columns 4, 5, and 6 give the values of the components in milliampere-centimeters. M_h is the component in the horizontal plane. It is seen that the order of magnitude of the average heart dipole must be about 1 ma-cm. To get the final value of M, M_h must be combined with M_z. Column 7 gives the angles of β, which is the angle between M_h and the $+X$-axis.

the different resistivities of the tissues.[1] A dipole was inserted in the heart area, and the distribution of potential around the thorax boundary was measured. In terms of distance along the boundary, Eqs. (8) can be expressed

$$M_x = k\int V\cos\theta dL,$$

$$M_y = k\int V\sin\theta dL. \tag{8A}$$

Also,

$$\tan\beta = M_y/M_x, \tag{13}$$

where θ is the angle between the normal to the boundary and the $+X$ axis, and β is the angle which the dipole makes with the X axis (Fig. 1). When the measurements were made in a homogeneous medium, the agreement between the known and calculated values of β was within 1°, and the calculated locations were correct to within 0.5 cm. In order to test the effect of electrical inhomogeneities in the medium, the measurements were repeated with insulating shapes representing the spine and sternum, and with lung sections having a conductivity equal to ¼ that of the surrounding medium. The surprising and gratifying result was obtained that the method still gave good results. The angles were all within 4°, and the locations within 0.6 cm, of the correct values. With completely insulating lung sections, the angles still agreed to within 4°, but the location was as much as 3 cm in error in the worst case. (The agreement might have been better in this last case if a lower value of k had been used to take into account the lowered average conductivity of the section.) There is also a close agreement between the magnitudes of the components in the homogeneous and inhomogeneous sections. Figures 2 and 3 show the equipotential field distributions in the two cases. Although there is a considerable distortion of the field in the neighborhood of the inhomogeneities, there is not a great deal of difference in the distribution of potential around the boundary of the sections. Thus the integrals (8A) give very similar results in the two cases. (The points marked "M" on the boundary are the points of maximum positive and negative potential.)

The method was also checked in a three-dimensional tank having the shape of an elliptical cylinder, 18 in. in depth. A dipole was located in the region corresponding to that of the heart in the body, i.e., nearer to the front and top of the tank. In a cylindrical volume conductor, with walls parallel to the Z axis, M_x and M_y are given by integrations over the walls only, and M_z by an integration over the top and bottom surfaces only. To obtain M_x and M_y, integrations of the type (8A) are performed around the tank wall at a number of horizontal levels, and the results plotted against Z. The area under the resulting curve gives the value of the double integral. The calculations were carried out for a known position of the dipole in the tank, and it was found that there was only 2° error between the measured and calculated values. This amount, however, was about the limit of error in adjusting the physical position of the dipole.

A similar procedure has been used to determine the horizontal component of the resultant dipole moment of the heart in the body (the heart vector). Potentials were measured completely around the thorax using up to 22 leads at 7 different horizontal levels. Curves were plotted of the distribution of potential around the thorax at each level, and the angle θ was measured for each point. Assuming a value of k representing the average conductivity of the thorax (1/500 mho-cm^{-1}), the values of M_x, M_y, and β were calculated for 6 instants of time during the RS complex of the electrocardiogram. The results are shown in Table I.

Although the method is very laborious to apply, it is thought to be more accurate than previous methods of finding the heart vector for the following reasons: (a) The finite boundary is taken into account. (b) No assumptions need be made as to the shape of the boundary. (c) The accuracy is not affected by the fact that the heart is near the front wall of the thorax, or that it is not small in size in relation to the thorax. (d) If one assumes a value of the average conductivity of the thorax, the magnitudes of the components are obtained with fair accuracy.

It is hoped that the method will prove useful in research, possibly for checking various systems of vectorcardiography.

[1] C. V. Nelson, J. Physiol. 116, 15P–16P (1952).

JOURNAL OF APPLIED PHYSICS VOLUME 25, NUMBER 4 APRIL, 1954

The Secondary Electron Resonance Mechanism of Low-Pressure High-Frequency Gas Breakdown*

ALBERT J. HATCH AND H. BARTEL WILLIAMS
Physical Science Laboratory and Physics Department, New Mexico College of Agriculture and Mechanic Arts, State College, New Mexico
(Received June 8, 1953)

Breakdown field strengths have been measured in air and hydrogen at pressures of the order of 1 micron Hg and frequencies from 25 to 90 Mc/sec between flat metal electrodes at separations from 1 to 4 cm. By suddenly applying a high voltage and then lowering it slowly an upper breakdown curve has been observed. This new curve can be combined with the lower breakdown curve and cut-off frequency such as reported previously by the Guttons, and by Gill and von Engel to form the boundaries of a breakdown region in the frequency-field strength domain.

The simple secondary electron resonance theories developed by Danielsson and Gill and von Engel have been extended by the semi-empirical determination of minimum and maximum limiting values of the electron-emission phase angle, a minimum limiting value of electron arrival energy, and a ratio between electron arrival and emission velocities.

I. INTRODUCTION

EARLY investigations of breakdown field strengths in gases at low pressures and high frequencies *ca* 1 to 100 Mc/sec were carried out by the Guttons[1] and by Kirschner,[2] using external and internal electrodes, respectively. Their principal observation was that the breakdown field strength decreased with decreasing frequency to values as low as 10 v/cm until a cut-off frequency was reached below which breakdown could be obtained only with high field strengths or not at all.

A later investigation by Backmark and Bengston[3,5] led to a theoretical analysis by Danielsson[4,5] in which he proposed a secondary-electron resonance breakdown mechanism. According to this mechanism, a few electrons initially present were accelerated to one electrode where they produced secondary electrons. These secondaries were emitted in a reversed electric field which carried them to the opposite electrode in approximately ½ cycle, producing another crop of secondaries which in turn were accelerated to the first electrode, and so on. If the secondary-emission yield was sufficiently greater than unity, then the number of oscillating electrons could be increased to a very large number in a short time. This large cloud of oscillating electrons caused sufficient ionization of the gas to result in breakdown of the gas.

The most recent work has been done by Gill and von Engel[6] who used external electrodes in the frequency range 12 to 75 Mc/sec. They observed that the breakdown field strength was independent of the type and pressure of the gas in the 1 to 8 micron Hg pressure range but that it was dependent on the secondary-emission characteristics of the type of glass of which the discharge tube was made. They also developed a theory postulated on the secondary-electron resonance mechanism which was somewhat different in its mathematical formulation from that of Danielsson and which included an important cut-off law relating the cut-off frequency to electrode separation.

This paper describes an extension of the previous work with internal electrodes.

II. EXPERIMENTAL APPARATUS AND PROCEDURE

A. Apparatus

A diagram of the electrode and discharge chamber assembly is shown in Fig. 1. For electrodes we have used flat, circular disks of aluminum, silver-plated copper, and copper, all with diameters of the order of 7 cm. The electrodes were supported inside a 20-cm internal diameter bell jar by means of a brass framework attached to a steel base plate. The large copper sheet mounted horizontally at the top of the brass framework and the grounded base plate provided an electrically symmetrical ground for the electrode system.

The rf voltmeter shown schematically in Fig. 2 is a modification of the peak-voltage measuring circuit used by Gill and von Engel. The special feature of this double diode peak reading voltmeter is that it is independent of asymmetry or unbalance in the rf system. The normal 30- to 120-v range of the Sensitive Research Corporation electrostatic voltmeter (E.S.V.M.) was extended to 1300 v by the use of bucking potentials provided by banks of hearing-aid batteries. A vacuum tube voltmeter (V.T.V.M.) was occasionally used in

* This work has been supported by the Bureau of Ordnance, U. S. Navy, through cooperation of the Applied Physics Laboratory of Johns Hopkins University, Silver Spring, Maryland.

[1] C. Gutton, Compt. rend. 178, 467 (1924); C. Gutton and H. Gutton, Compt. rend. 186, 303 (1928); H. Gutton, Ann. phys. 13, 62 (1930).

[2] F. Kirschner, Ann. Physik 77, 287 (1925); Ann. Physik 7, 798 (1930).

[3] N. E. Backmark and U. Bengston, Diplomarbeit K. Techn. Hochschule Stockholm (1941).

[4] U. Danielsson, Diplomarbeit K. Techn. Hochschule Stockholm (1943).

[5] H. Alfvén and H. J. Cohn-Peters, Arkiv för Matematik, Astronomi och Fysik 31, 1 (1944). This contains a summary in German of the work of Backmark and Bengston and of Danielsson (see references 3 and 4) and others.

[6] E. W. B. Gill and A. von Engel, Proc. Roy. Soc. (London) A192, 446 (1948).

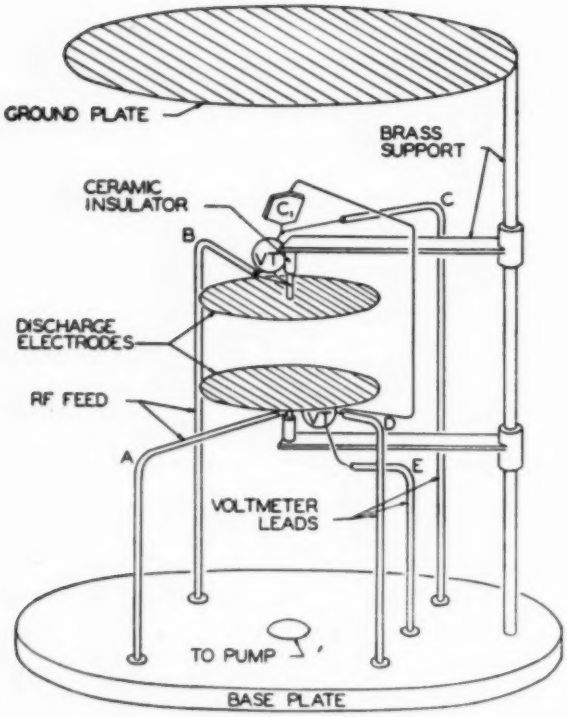


FIG. 1. Arrangement of discharge electrodes and electrical components inside bell jar.

place of the E.S.V.M. Standard high-frequency voltmeter calibration methods were used. Electrode separations were measured with a vernier caliper to within 0.05 cm. The over-all uncertainty in the breakdown field strength values varied from approximately ± 2 percent to ± 10 percent, as specified in Sec. III.

Conventional circuits were used for the rf power supply. Frequency measurements were made with a Measurements Corporation megacycle meter to within $\frac{1}{4}$ Mc/sec. A two-stage silicone-oil diffusion pump was used in the vacuum system. Pressures were measured directly with a D.P.I. Pirani gauge which was calibrated periodically against a triple-range McLeod gauge in which Dow-Corning DC-702 silicone oil was used. The bell jar was sealed to the base plate with Apiezon sealing compound. Gases were admitted through a controlled leak and drying tube. The vacuum line from the bell jar to the diffusion pump was not

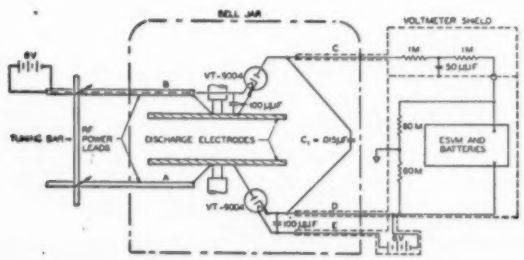


FIG. 2. Schematic diagram of voltmeter circuit showing connections to discharge electrodes.

trapped. The system was kept mercury-free, but no attempt was made to maintain gas purity since it has been shown that the nature of the gas is of relatively minor importance in this type of breakdown. For this reason the indicated use of air, oxygen, and hydrogen in this work does not connote gas purity in the usual sense.

B. Procedure

Measurements of lower breakdown voltage were made by slowly increasing the plate voltage of the 832-A rf amplifier and observing the electrostatic voltmeter reading. At the instant breakdown occurred the voltmeter reading would start to decrease. The highest voltage observed just at this instant of breakdown was recorded as breakdown voltage. Breakdown was usually accompanied by a blue glow between the electrodes.

In order to make measurements of the upper breakdown voltage, it was first necessary to raise the voltage above the breakdown region. This was done by either of two methods. The first method was to apply full plate supply potential suddenly by switching on the final rf amplifier. The second method was to short the bell jar resonant circuit, apply full plate supply potential, then break the short circuit. In either case the rf voltage rise through the breakdown region was too rapid to cause breakdown. The rf voltage was then lowered slowly by lowering the rf amplifier plate supply potential. When breakdown occurred, the electrostatic voltmeter reading would decrease very rapidly and a blue glow would appear between the electrodes. The lowest voltage observed just before this sudden drop in voltage occurred was recorded as the upper breakdown voltage. We call this the "jumping" technique.

III. EXPERIMENTAL RESULTS

A. Upper and Lower Breakdown Curves

Results of measurements of breakdown field strength *vs* frequency for 7.5-cm diameter machined alclad electrodes at a separation of 3 cm in dry oxygen are shown in the log-log plot in Fig. 3. The four different sets of data were obtained on different days within a three-week period. The V.T.V.M. was used in place of the E.S.V.M. in these measurements. The over-all uncertainty in the values of field strength for these data is about ± 10 percent below 70 Mc. The rather large spread of lower breakdown values above 70 Mc/sec was believed to be the result of resonances in the voltmeter circuit. Reproducibility of the data below 70 Mc/sec is fairly well established.

The lower breakdown curve confirms the previously reported existence of very low breakdown field strengths and a cut-off frequency. It was impossible to obtain breakdown below 34 Mc/sec with field strengths as high as 300 v/cm. The outstanding feature of these data is the upper breakdown curve from 34 to 58 Mc/sec. Within this frequency range the upper and

lower breakdown curves define the shaded region within which breakdown can occur. Outside this region we were unable to obtain breakdown. Because of rf voltage limitations it was not possible to obtain measurements of the upper breakdown curve above 58 Mc/sec.

B. Closed Breakdown Region

In Fig. 4 are shown the results of measurements made in hydrogen to determine whether or not a closure of the upper and lower breakdown curves could be observed in the cut-off region. The E.S.V.M. was used in these measurements and the values of field strength are accurate to within approximately ±3 percent. Stability of the electrode surface conditions was achieved by burning the discharge for several hours prior to making measurements. The glow was very faint at the 0.10-micron pressure used here. Breakdown was first obtained on the lower curve at 81 Mc/sec, and subsequent measurements were made in order of decreasing frequency as indicated by the arrow. Breakdown could not be obtained below 27.5 Mc/sec. At this stage in the experiment the measurements were taken in the order of increasing frequency, as indicated by the arrow, and the procedure for obtaining the upper curve as described in Sec. II B was used. In the immediate vicinity of cutoff, breakdown was evidenced only by the appearance of a faint glow; there was no drop in rf voltage. The continuity of the observed breakdown field strengths in the cut-off vicinity appears to indicate the existence of a definite observable closure of the shaded breakdown region.

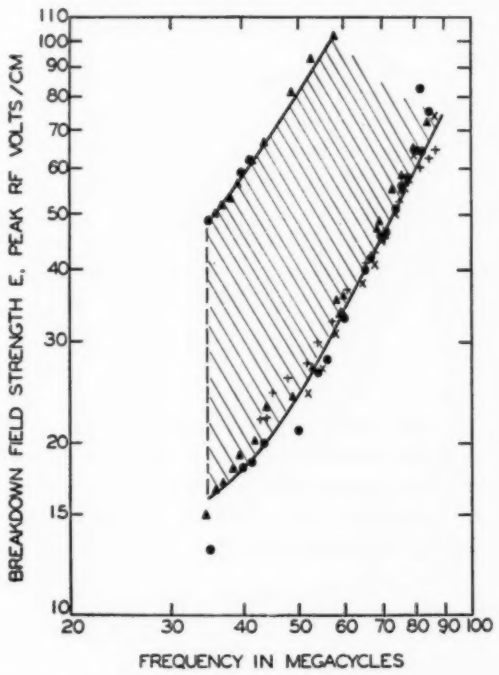


FIG. 3. Log-log plot of lower and upper breakdown field strengths for alclad electrodes. Electrode separation: 3 cm; electrode diameter: 7.5 cm; atmosphere: dry oxygen at less than 1 micron Hg pressure; data taken on different days in the order ●, +, ×, ▲; broken line indicates cut-off frequency. The breakdown region is shaded.

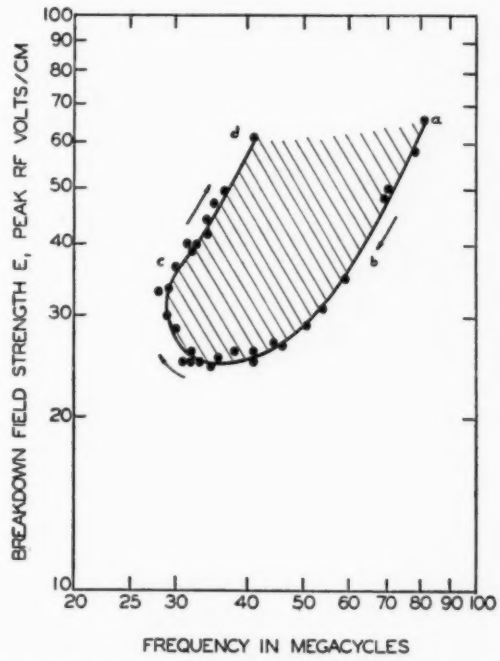


FIG. 4. Log-log plot of closed breakdown curve for silver-copper electrodes. Electrode separation: 3 cm; electrode diameter: 6.5 cm; atmosphere: hydrogen at less than 0.10 micron Hg pressure; data taken in order indicated by arrows. The breakdown region is shaded.

C. Cut-Off Frequencies

The results of measurements to confirm the general relation between cut-off frequency and electrode separation in air are displayed in Fig. 5. Measurements of lower breakdown voltage were made in the order of decreasing frequency for each electrode separation until the cut-off frequency was reached. Values of voltage for these data are accurate to within approximately ±2 percent.

Reproducibility of these data was checked by repeating the 3-cm electrode separation measurements after an interval of 10 days. During this time the bell jar had been opened five times to change the electrode separation, and approximately 60 breakdown measurements had been made. Cutoff was observed between 31 and 32 Mc/sec in both cases. The observed values of breakdown voltage for the two cases agreed within 2 percent except in the immediate vicinity of cutoff, where the variation was as great as 20 percent. This latter variation was attributed to the high sensitivity of breakdown voltage to changing electrode-surface conditions near cutoff.

D. Electrode-Surface Conditions

Dependence of breakdown on the secondary-emission characteristics of the electrode surfaces was observed

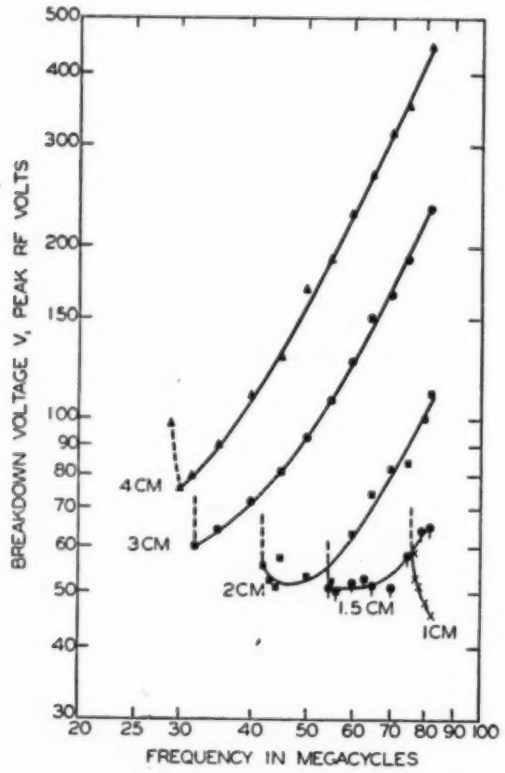


FIG. 5. Log-log plot of breakdown voltage as a function of indicated electrode separation for copper electrodes. Electrode diameter: 7.6 cm; atmosphere: air at 1 micron Hg pressure; broken lines indicate cut-off frequencies.

with an appropriately modified apparatus using alclad electrodes at a separation of 3.5 cm and a frequency of 50 Mc/sec. With gas pressure reduced to 0.2 micron Hg, the secondary-electron yield δ was observed to be 1 for a primary-electron energy of 25 ev, and breakdown occurred at a field strength of 31 rf v/cm. After outgassing the electrodes by burning the discharge for several minutes with pressure held constant, it was found that $\delta=1$ for a primary-electron energy of 53 ev and that breakdown occurred at a field strength of 42 rf v/cm. Changes in breakdown field strength were avoided in this work either by keeping the electrodes gassy at 1 micron Hg pressure as for the data in Figs. 3 and 5 or by outgassing the electrodes to the system limit as for the data in Fig. 4. Prolonged burning of the discharge by several weeks of use produced colored films on the metal electrodes.

IV. THEORY

A. The Significance of Resonance

A qualitative description of the secondary-electron resonance breakdown mechanism has been presented in Sec. I. The term "resonance," however, is used here in a different sense from that ordinarily associated with simple harmonic motion. At the start of a half-cycle transit period most of the electrons leave the emitting electrode surface with energies of the order of 5 ev. At the end of this transit period most of these electrons arrive at the opposite electrode with energies of the order of 50 to 500 ev. It is the cyclic repetition of this process which we refer to as resonance.

B. Assumptions

The assumptions as stated here are essentially the same as those used by Danielsson and by Gill and von Engel.

The fundamental hypothesis for this theory is the existence of the secondary-electron resonance mechanism. In order that this mechanism be operative it is necessary to assume that the electron mean free path and the wavelength of the applied high-frequency field are both large compared to the electrode separation.[7] For mathematical simplicity it is convenient to assume that all electrons have $\frac{1}{2}$-cycle transit times. In conformity with the treatment of Gill and von Engel, it is assumed that the ratio of the electron arrival to emission velocity is a constant $k=v/v_0$ [8] and that secondary-electron emission velocities are normal to the electrode surfaces. It is further assumed that the electric field between electrodes is space-uniform, that space charge effects are negligible, that electron-arrival energies exceed the ionization potential of the gas, and that a few electrons are produced randomly between the electrodes by natural processes. The extent to which many of these assumptions compensate in an undetermined manner for other processes in this breakdown mechanism is not known.

C. Resonance Equations

Although the standard electron equation of motion on which the simple theory is based was used by both Danielsson, and Gill and von Engel, the major part of the following analysis is due to the latter.

For an electron in a sinusoidal field of peak strength E and frequency $f=\omega/2\pi$, the motion equation is

$$m\frac{d^2x}{dt^2}=eE\sin(\omega t+\phi), \tag{1}$$

where ϕ is the secondary-emission time phase angle. There is no frictional or collision term in (1) by virtue of the mean free path assumption. The first integral of (1) evaluated for $\frac{1}{2}$-cycle transit time yields the expression for arrival velocity

$$v=v_0+\frac{2eE\cos\phi}{m\omega}, \tag{2}$$

[7] The kinetic theory electron mean free path in H_2 at 0°C and 1 micron Hg pressure is approximately 80 cm; for N_2 it is approximately 40 cm.

[8] This assumption is not an accurate representation of secondary-emission characteristics but is very useful in getting a simple formulation of the problem and leads to good correlation between theory and observation.

where v_0 is the emission velocity. The second integral of (1) evaluated for ½-cycle transit time, and a trajectory equal to the electrode separation d leads to the transit equation

$$d=\frac{v_0\pi}{\omega}+\frac{eE(\pi\cos\phi+2\sin\phi)}{m\omega^2}. \tag{3}$$

This can be written more conveniently for our purpose as

$$E=\frac{\omega^2 d-v_0\pi\omega}{(e/m)(\pi\cos\phi+2\sin\phi)}. \tag{4}$$

Application of the Gill and von Engel assumption $v/v_0=k$ gives (2) and (4) as

$$v=\left(\frac{k}{k-1}\right)\frac{2eE\cos\phi}{m\omega}, \tag{5}$$

and

$$E=\frac{\omega^2 d}{(e/m)\Phi}, \tag{6}$$

respectively, where

$$\Phi=\frac{k+1}{k-1}\pi\cos\phi+2\sin\phi. \tag{7}$$

Gill and von Engel continued their analysis by solving Eqs. (5) and (6) simultaneously, using ϕ as a parameter and fitting to the data at $\phi=0°$.

At this point we add two observations. First, the electron-arrival energy can be expressed as

$$eW_f=\tfrac{1}{2}mv^2, \tag{8}$$

which can be combined with (5) and (6) to give the frequency f as

$$f=\frac{(k-1)\Phi}{k\pi d\cos\phi}\left(\frac{eW_f}{8m}\right)^{\frac{1}{2}}. \tag{9}$$

Second, at fixed values of ω and d the minimum E in (6) occurs for maximum Φ. Maximizing Φ with respect to ϕ gives the condition

$$\phi=\tan^{-1}\left(\frac{k-1}{k+1}\cdot\frac{2}{\pi}\right). \tag{10}$$

This procedure yields a group of semi-empirical relations, Eqs. (6), (7), (9), and (10), which can be used in calculating the breakdown region as in the illustration which follows.

D. Calculation of the Breakdown Region and Boundary Values

In this illustration the theory will be fitted to the data in Fig. 4 for silver-plated copper electrodes at a separation $d=3$ cm. Equation (6) fits the straight line segment a–b of the lower breakdown curve in Fig. 4 when $\Phi=6.6$. Use of this value in (7) with (10) gives $k=3$, $\phi=+18°$. If k is held constant, ϕ can be used as a parameter in (6) to obtain the family of straight lines of slope 2 shown in the log-log plot of Fig. 6. The resonance line A–B for $\phi=+18°$ in Fig. 6 corresponds to a–b in Fig. 4. In this manner (6) is found to fit the straight line segment c–d of the upper breakdown curve in Fig. 4 when $\phi=-56°$. The resonance line C–D in Fig. 6 corresponds to c–d in Fig. 4. From (9) a point on each of this family of straight lines in Fig. 6 can be found corresponding to a particular electron-arrival energy W_f. For an arrival energy of 60 ev this gives a constant arrival energy locus B–C as shown. The theoretical breakdown region shown shaded in Fig. 6 is A–B–C–D.

The boundary values determined in this illustration are $k=3$, $-56°\leq\phi\leq+18°$, and $W_f=60$ ev. By way of comparison, Danielsson assumed $k=\infty$, i.e., $v_0=0$, $0°\leq\phi\leq90°$, and found $W_f=80$ ev for aluminum electrodes at various separations and a frequency of 65 Mc/sec. Henneburg, Orthuber, and Steudel[9] obtained the theoretical limits $0°\leq\phi\leq32.5°$ for the case of zero electron-emission energy. Gill and von Engel found, by fitting their equations to data for external electrodes at a separation of 3 cm, that $k=4$, $W_f=90$ ev, and that cutoff occurred at an escape limiting value of $\phi=-58°$. Their treatment did not incorporate a second limiting value of ϕ. These values of k of the

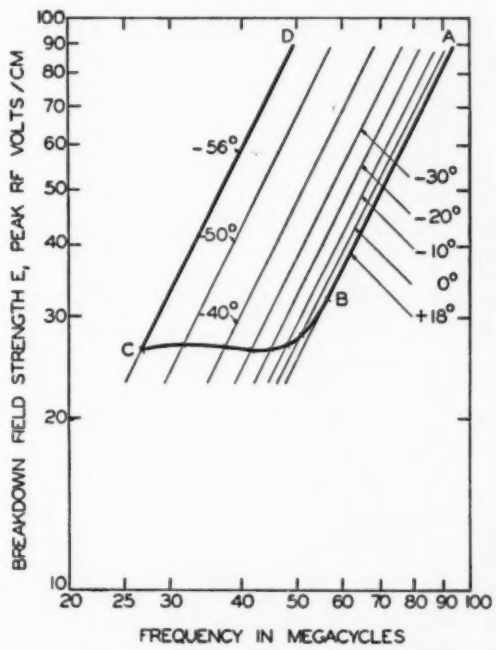


FIG. 6. Log-log plot to illustrate theoretical calculation of breakdown region for 3 cm electrode separation, minimum electron-arrival energy of 60 ev, and $k=3$. Light lines represent conditions for ½-cycle electron transit time at indicated phase angles. Heavy curve represents boundary of theoretical breakdown region.

[9] Henneburg, Orthuber, and Steudel, Z. tech. Phys. 17, 115 (1936).

order of 3 to 4 and electron-arrival energies of the order of 60 to 90 ev are compatible with known secondary-electron emission energies and yield.

E. Cut-Off Frequencies and Voltage

The cut-off frequency f_{co} at point C in Fig. 6 can be obtained by using the upper breakdown phase angle boundary value ϕ_u in (9),

$$f_{co}=\frac{(k-1)\left(\frac{k+1}{k-1}\pi\cos\phi_u+2\sin\phi_u\right)}{k\pi d\cos\phi_u}\left(\frac{eW_f}{8m}\right)^{\frac{1}{2}}. \quad (11)$$

For constant k, ϕ_u, and W_f this becomes

$$f_{co}=\frac{C_1}{d}, \quad (12)$$

where C_1 is determined by fitting. This is the equivalent of the cut-off law obtained by Gill and von Engel from a general similarity theorem.

The breakdown *voltage* at cutoff can be obtained as

$$V_{co}=\frac{W_f(k-1)^2\left(\frac{k+1}{k-1}\pi\cos\phi_u+2\sin\phi_u\right)}{2k^2\cos^2\phi_u}. \quad (13)$$

This relation is independent of electrode separation and applied frequency, a fact which is useful in studying electrode geometry effects. For fixed k and ϕ_u this becomes

$$V_{co}=C_2W_f, \quad (14)$$

where C_2 can be determined by fitting.

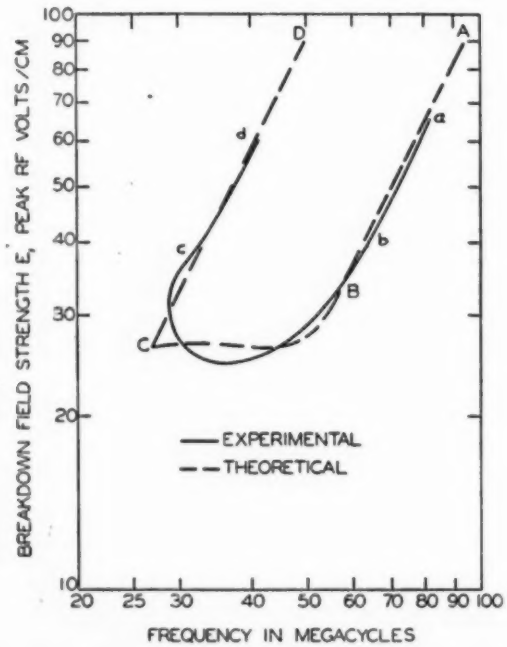


FIG. 7. Comparison of experimental closed curve from Fig. 4 and fitted theoretical curve from Fig. 6.

V. COMPARISON OF EXPERIMENTAL RESULTS WITH THEORY

A. Breakdown Region

The experimental breakdown curve from Fig. 4 and the theoretical curve from Fig. 6 are plotted together in Fig. 7.

Because of the fitting process used in Section IV D the comparisons which can be made between theory and observation are limited to features other than the determination of boundary values. The best correlation is between the slopes[10] of the resonance portion of the lower breakdown curve A–B. The fact that this slope is approximately 2 in the log-log plots of all measurements within error limits is a strong justification of the invariance of Φ along the lower breakdown curve and of the relation between k and ϕ as expressed in (7). The correlation between theoretical and experimental slopes of the resonance portion of the upper breakdown curve C–D is very nearly as good as for the lower breakdown curve. The correlation in the energy limit locus region B–C, especially near C, the cut-off point is not very good. The theoretical concept of a wedge-shaped corner of the breakdown region between B–C and C–D appears to have no observable physical counterpart. This indicates the following inadequacy in the theory: The simplifications in this theory are representations of the dominant processes of phase angle and energy limitations in the secondary-electron resonance multiplication mechanism. In the transition zones where the energy limitation replaces the phase-angle limitation as the dominant process the simple theory indicates an abrupt transition, whereas it actually appears to be a gradual one.

B. Cut-Off Frequencies and Voltage

When the theory is fitted to the breakdown data for electrode separations 1 to 4 cm as shown in Fig. 5, the approximate lower boundary values found are $\phi=+22°$, $k=4$, $W_f=40$ ev. Upper breakdown values were not obtained in this case; hence, the negative limiting ϕ cannot be determined as before. However, assuming constant k, the approximate f_{co} for the 1.5-cm separation can be obtained by using $\phi_u=-40°$. If these boundary values are used in (11), the cut-off law (12) becomes $f_{co}=79/d$, from which the theoretical cut-off frequencies in Table I are calculated. The fact that at the larger separations the observed f_{co} is considerably greater than the calculated f_{co} is apparently due to increased electron loss to the sides. The experimental

[10] The principal advantage of the log-log plots in this work lies in the comparison of the theoretical and experimental slopes of the resonance portion of the breakdown curves. The onset of the energy limit also is distinguished clearly.

values for f_{co} were all obtained for an electrode diameter of 7.6 cm.

In Fig. 8 are plotted the cut-off frequencies observed in the present work along with similar values obtained by the Guttons and by Gill and von Engel. These values of f_{co} are plotted as a function of the reciprocal of the electrode separation $1/d$. The theoretical relation from (12) is also plotted for comparison. The details of the experimental conditions for the work of the Guttons and of Gill and von Engel are different from those in the present work. The correlation between the cut-off law as expressed in (12) and in the observed f_{co} values over an electrode separation range of 1 to 20 cm as shown in Fig. 8 appears very promising.

According to (13) the breakdown voltage at cutoff is independent of electrode separation. This independence can be shown to apply to the entire energy limit locus *B–C*. If no electrode geometry effects were present, one would expect to find identical energy limit loci. This appears to be approximately the case for the 1.5- and 2-cm separations in Fig. 5. For the 3- and 4-cm separations, however, the values of V increase, apparently because of the poorer electrode geometry.

TABLE I. Cut-off frequency f_{co} for various electrode separations.

Electrode separation d in cm	Theoretical f_{co} in Mc/sec	Observed f_{co} in Mc/sec
1.0	79	76±2
1.5	53	54±2
2.0	40	42±2
3.0	26	31±2
4.0	20	28±2

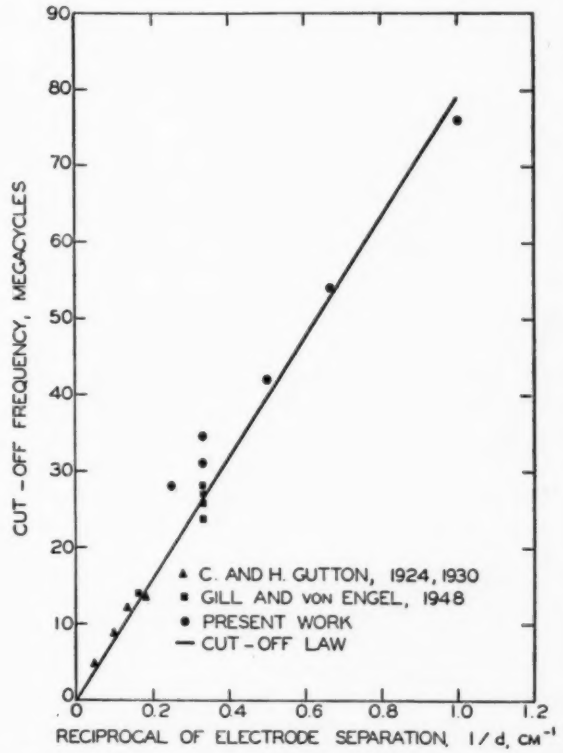


FIG. 8. Summary of cut-off frequencies *versus* reciprocal of electrode separation from data of several observers. All values from previous observers for external electrodes. Present values for internal electrodes. Cut-off law plotted from $f_{co}=79/d$, f_{co} in Mc/sec, d in cm.

VI. DISCUSSION

The observation of a closed breakdown region and the correlation with the modification of the secondary-electron resonance theory substantiate and clarify the secondary-electron resonance mechanism of this species of breakdown. Additional confirmation of the existence of this mechanism has been obtained in the measurement of electron-arrival energies.[11] As an example, for a 2-cm electrode separation at $E=67$ v/cm and $f=70$ Mc/sec we have $\Phi=3.27$. We can use the previously obtained $k=4$ for Fig. 5 and find $\phi=-33°$ which can be used in (5) to give an arrival energy of 102 ev. This value is in good agreement with the observed peak of this arrival-energy spectrum between 85 and 115 ev and suggests that a sort of electron bunching occurs. Here, bunching means a limitation in the range of ϕ for which resonance is most likely for specified parameters. For example, referring to Fig. 6, along the resonance line *A–B* the bunching will be about $\phi=+18°$; as field strength is increased at a fixed frequency, the bunching will shift first to 0° then to negative values of ϕ. Electrons emitted within a phase angle range of, say, $\pm 10°$ will traverse the interelectrode space in a fairly well-defined group and will reach the opposite electrode with the observed range of arrival energies.

Another confirming observation has been that of electron resonance without any visible discharge or attenuation of applied voltage at pressures of the order of 0.05 micron Hg. For a fixed frequency and electrode separation, electron resonance has been observed continuously within a field-strength range approximately the same as that within which visible breakdown occurs at high pressures.

We take pleasure in thanking Dr. G. W. Gardiner, Dr. R. W. Dressel, and Mr. J. S. Arnold of the Physical Science Laboratory for valuable discussions of this work and for critical reading of the manuscript. We also are indebted to Dr. S. N. Foner of the Applied Physics Laboratory for valuable suggestions concerning this paper and to Dr. Sigurd Johansen of New Mexico College of Agriculture and Mechanic Arts for assistance in translating the Swedish references 3 and 4.

[11] H. B. Williams and A. J. Hatch, Phys. Rev. 89, 339 (1953).

Switching in Bistable Circuits

R. STUART MACKAY
Division of Electrical Engineering, University of California, Berkeley, California
(Received August 5, 1953)

Switching or triggering in certain nonlinear circuits having two stable states is studied in graphic detail. The minimum and maximum pulse requirements are discussed and some of the general ideas of bistability considered. A convenient technique for studying the response of circuits, apparently in slow motion, is described.

ALL triggering processes seem capable of description in terms of a curve with a region of negative slope. Those involved in switching and counting circuits are of especial interest in our modern technology. The first few of the following sections considers the action of a device that allows one to follow visually the details of the triggering mechanisms. A number of the concepts that often seem a bit artificial are clearly demonstrated.

The term triggering is most properly applied to those situations in mechanics, chemistry, physiology, physics, electricity, etc., in which there are two possible states, the intermediate ones being excluded or unstable. The situation can be described in terms of a hill or potential barrier or, in the electrical case, it is found expedient to describe switching in terms of transitions between opposite sides of a region displaying a negative slope on the E–I curve (sometimes called a region of negative resistance). Most counters employ the class of negative resistance in which voltage is a multiple-valued function of current, though they can be built from the type in which current is a multiple-valued function of voltage.[1] To achieve bistability with the former class, one needs a series resistance greater than the magnitude of the negative resistance, while to trace out its properties the total series resistance must be less than the magnitude of the negative resistance. The converse is the case for the other class of elements. The following device was set up to display a transition between states in one of these tetrode-like negative resistances.

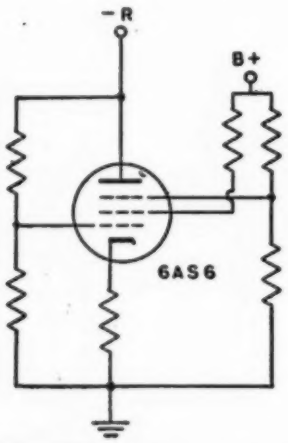


FIG. 1. A stable negative resistance having the form of the plate characteristics of a tetrode (as opposed to that near the knee in the plate curves of a beam power tube).

[1] R. S. Mackay, J. Appl. Phys. 24, 1163 (1953).

THE NEGATIVE RESISTANCE

It was felt desirable that the negative resistance be stable but adjustable and that it display its properties with respect to ground; thus neither a tetrode nor a multivibrator was used. One was constructed using the scheme shown in Fig. 1. Here an increase in plate voltage, by virtue of the grid resistors, causes an increase in control grid voltage. This causes the cathode potential to rise by the usual cathode follower action. The suppressor grid, being fixed in potential, eventually goes negative with respect to the cathode and thereby the plate current is stopped (goes to the screen). As one continues to increase the plate voltage, the current will again begin to increase because of the current drawn through the control grid resistors. Thus, a plot of current as a function of voltage is an "N" curve as in Fig. 2. One can readily control the shape of this curve by changing the size of the cathode resistor. One can alter its rate of fall by controlling the ratio of the two grid resistors (one can make it more rapid by replacing the top one by a neon tube). The rate of rise of the right-hand portion can be increased by decreasing the size of the grid resistors, while keeping their ratio fixed.

DISCONTINUOUS TRANSITION TRACING

If one takes the previous curve and draws upon it a sufficiently flat load line there can be three intersections of which the outer two represent stable conditions of current and voltage in the negative resistance. If the circuit is in the left-hand state, an increase in voltage will cause it to switch to the right-hand state provided this increase is sustained for greater than a critical length of time (Fig. 2). The following device makes clear just what determines this length of time, i.e., at what point after the removal of a switching pulse such a unit will be regeneratively carried to the second state rather than back towards the state from which it came. During the transition caused by translating the load-line parallel to itself to the right, one observes the transition plotted in terms of instantaneous current as a function of instantaneous voltage. This transition is displayed superimposed upon the negative resistance curve itself. In Fig. 3 is seen the wave form applied to the circuit under test. One first traces the negative resistance curve from its far end toward the origin and

then unshorts the series resistor required for bistability with the help of a relay (Western Electric type 275C). An increase in the supply voltage back up from zero thus leaves the circuit in its left-hand state. Once there, a rectangular positive pulse is superimposed upon the supply voltage, thus starting the switching process towards the right-hand state. After a sufficient time the series resistor is again shorted out preparatory to the tracing of another negative resistance curve. Rapidly repeated, this process seems to display the transition superimposed upon the characteristic curve. A number of such traces are displayed later in this paper.

TIME AMPLIFIER OR SLOW-MOTION CONTROL

The previous curves give no idea of the rapidity with which a transition takes place, or more important, the relative speed with which different portions of the curve are traversed. It would seem desirable to slow

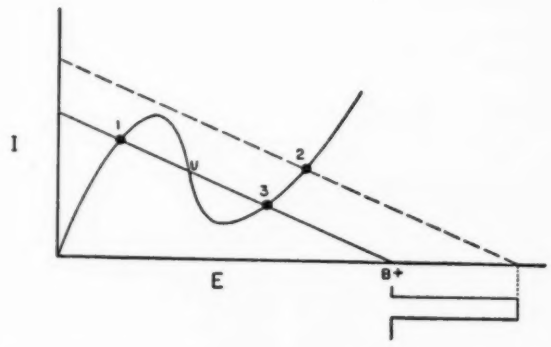


FIG. 2. Negative resistance curve and load lines. Momentarily increasing the voltage applied to this series combination of positive and negative resistance while in state 1 will force it to go toward state 2, after which it may go to state 3; or it may return to state 1 if it has not gotten beyond the unstable point U.

the action without introducing abnormal conditions, i.e., without artificially loading the whole circuit with capacity. One can apparently slow down the transition to an arbitrary degree by suitably modulating the intensity of the oscilloscope tracing. In particular, if one applies to the cathode of the cathode-ray tube a square pulse of linearly increasing width, then slightly more of each successive transition will appear and the whole transition will seem to require a time given by the period of the sawtooth wave modulating the square-wave generator. For example, if the basic pattern is being repeated 20 times per second and on the first tracing one displays 1/60 of the pattern, 2/60 on the second tracing, 3/60 on the third, etc., then a full transition will apparently require three seconds. Relative velocities will be maintained correct in all parts of the slow-motion picture. If, rather than seeing the pattern gradually unfold, one would prefer to see a spot moving along the pattern with proper relative speed, one need merely apply the derivative of the

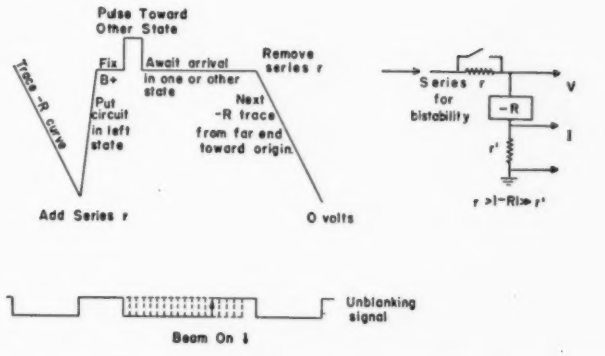


FIG. 3. Wave forms applied to a circuit to display the path of the operating point during a transition between states. By suitable periodic sampling of the condition of the circuit (intensity modulation of the oscilloscope beam) one obtains a slow motion representation.

above-mentioned signal to the cathode of the oscilloscope tube (or generate a narrow unblanking pulse at the end of the previously mentioned rectangle; its width then determines the effective screen persistence). This cyclic linear modulation can be replaced by a calibrated manual control to take precise data on the rate of progression.

A diagram of the whole circuit is given in Fig. 4. The basic wave form of Fig. 3, that is applied to the series combination of resistance and negative resistance, is generated by a Miller transitron upon whose plate wave shape is superimposed, at a suitably delayed time, the switching pulse. The "time expander" pulses are generated by a phantastron that is slowly duration-modulated through increasing times by a bootstrap circuit which sets the initial plate potential from which "run-down" starts, and scans it through linearly increasing values. This long period bootstrap circuit uses a small neon tube in place of one of the usual pair of isolated voltage sources. It is the slope or period of its triangular wave that determines the apparent slowing down or duration of the displayed transition.

RESULTS

The photographs of Fig. 5 display typical observed patterns. It is seen that during a pulse the operating point tends to travel along the load line (rather than following the negative resistance curve or traversing some intermediate path) in the particular unit as thus far described. If the pulse ends before the circuit has had time to reach the center (unstable) intersection of the load line and negative resistance curve, the circuit will be carried back to the original intersection from which it came. If the pulse lasts long enough for the circuit to progress beyond this center intersection then regeneration will carry the circuit forward into the new state. If the pulse lasts longer than the time required for a full transition, then the circuit will wait for the end of the pulse in the state determined by the intersection of the temporary load line and an upper

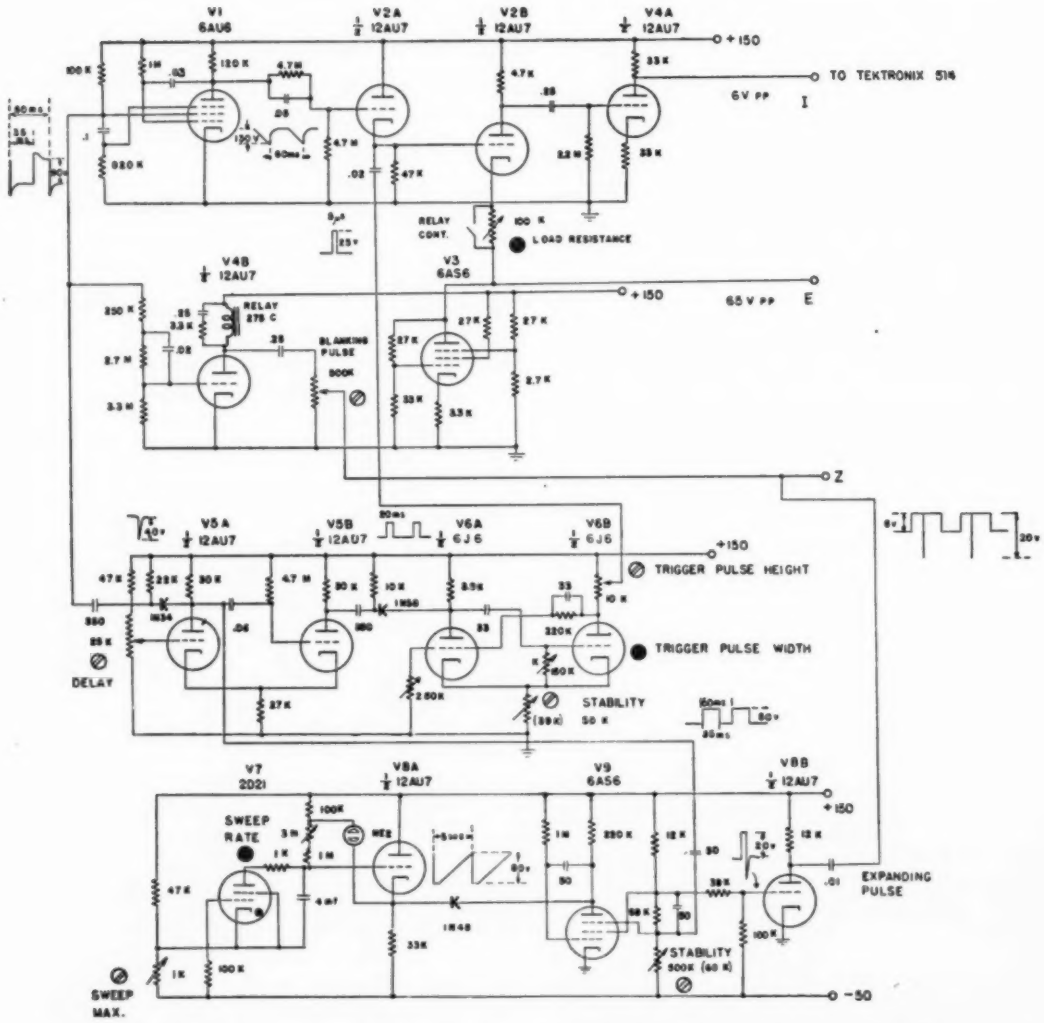


FIG. 4. Detailed circuit of transition plotter. For a greater repetition rate but larger capacitive loading, the relay can be replaced by a tube whose grid is swung between B^+ and ground through a several megohm resistor; or a double diode clamp can be activated through a transformer to short the resistor periodically.

right portion of the negative resistance curve, after which a transition will return it to the right-hand intersection of the normal load line and the curve.

If the load resistance is lowered until the load line is steep enough to intersect the curve in only one point, then a pulse will produce the effect shown in the upper part of Fig. 6. Similarly, with a small pulse height and flat load line the circuit will switch between states all to the left. It is also possible to build circuits that are in one state at one voltage level and a different state at another voltage level. In the lowest picture is shown a situation in which the circuit switches to the right-hand state during the pulse but returns to the left-hand state after the original voltage level is returned.

Some of these pictures, by variable trace density, hint at the slowing down of the transition in passing over and breaking away from the left-hand peak of the negative resistance curve. This effect is of course very noticeable when watching the curve being traced out in slow motion. When using this time modulation, one can also study the effects of pulse height (as well as pulse width) on transition speed. With the help of this device one can, for example, determine the optimum pulse for producing a transition in the minimum possible time. This question is somewhat reminiscent of the classic brachistachrone problem.

In Fig. 7 is shown the transition that takes place with a different distribution in the stray capacities. This photograph was taken by loading the series resistor with 75 picafarads. Because of the tendency for such capacity to temporarily hold the voltage across the involved resistor fixed, there results overshoots and a transition that follows more closely the curve shape of the other element. It is the stray capacities, whose effects are not included in such a load-line diagram, that supply or take up any unaccounted for currents involved in a transition. In fact, a continuous transition is not possible without stray capacities; if it

were not for them, there would be no nonforbidden path between states that did not violate one of the constraints imposed upon the circuit by the two major elements being in series.

Not only does the foregoing demonstrate that the whole concept of the role of negative resistance is not merely an expedient fiction, but it also suggests, for example, why one might expect to be able suitably to pulse a nerve for a sufficiently short time without setting up a nervous impulse.[2] It is similarly obvious that one can push a "microswitch" for a short enough time to give no output, just as one can momentarily overvoltage a discharge tube without breakdown. In each case it is clear why a temporarily increased sensitivity follows. Even the married state with its unstable intermediates and regeneratively mounting emotions might be discussed here.

The preceding also gives us a better understanding of such things as the plate wave form of a monostable multivibrator whose triggering is shown in Fig. 8 (a 1-shot multivibrator was used, rather than a bistable one, merely for convenience; the two are quite equivalent in the time range shown, which was much less than the duration of the abnormal state). This figure shows the rise of a plate when the circuit is triggered by several different pulses, all being in the vicinity of one just barely long and high enough. In some cases, the circuit can be seen to sit at an intermediate state for an appreciable time before gradually returning to the original state. This corresponds to triggering the circuit very close to that central condition beyond which it would be carried into the other (in this case abnormal) state. The shape of these curves can be inferred from the other type of observation.

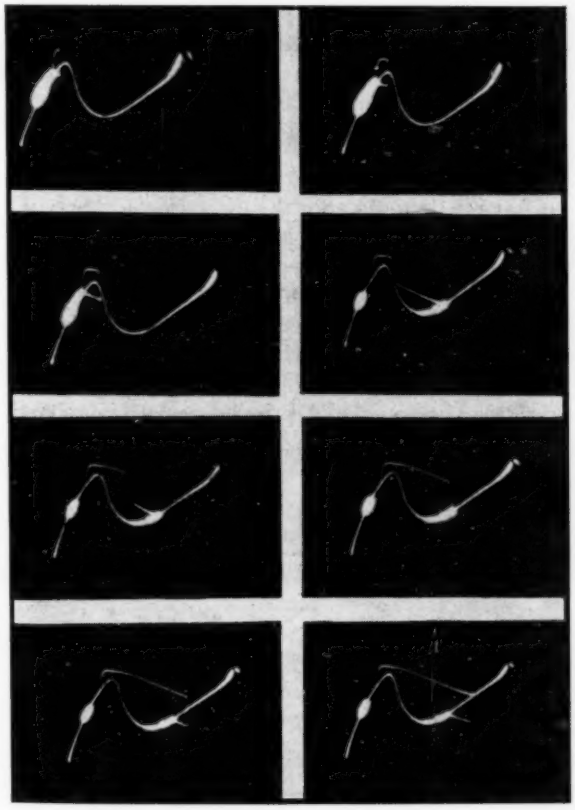


FIG. 5. The transitions observed with successively increasing switching-pulse widths. A similar progression is observed if it is the pulse height that is varied.

[2] R. S. Mackay and S. E. Leeds, J. Appl. Physiol. 6, 67 (1953).

This photograph of the response of a one-shot multivibrator to short pulses of varying height is reminiscent of the action of pushing a ball up the side of a hill. Sometimes it will coast over to the other side, and sometimes back; sometimes it will perch for a microsecond "on the crest" before going one way or the other. (The crest of the hill corresponds to the central intersection in the $-R$ picture). In all cases it should be remembered that the transition to the other state was much shorter than the duration of the other state. The integral of a function of the pulse height with respect to time seems to determine the impulse given the circuit, the height determining the speed and the pulse duration how far one gets.

In all cases of bistability the output quantity affects the input point, and thus it is appropriate to speak of the loop gain in this feedback system. A loop gain of at least unity must be achieved to accomplish switching through the normally excluded region. As seen above, cutoff and grid current, or limitation of swing by clamping, are not necessary to quiescence though they protect against disturbance in sensitive circuits. The loop gain and speed of progression seem to increase with the vertical separation between the "N" curve and the momentary operating point.

For comparable widths, a shallow $-R$ or low hysteresis indicates a low loop gain that will only slowly

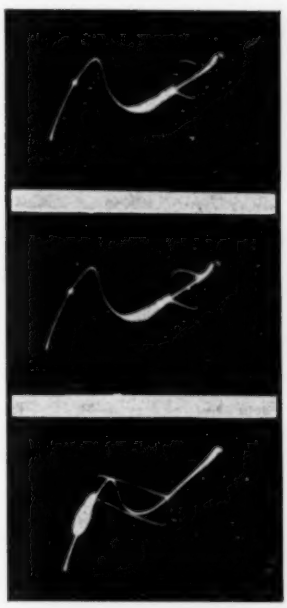


FIG. 6. Pictures taken with different steady-state conditions. The top two are with a small load resistance, and the bottom one is with a large load resistance. The top ones involve essentially r and c since the $-R$ tube is inactive.

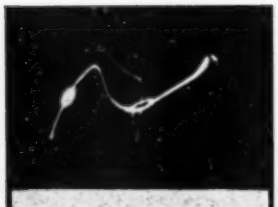


FIG. 7. Transitions with a different distribution of associated capacities.

drive a circuit against retarding effects. The measurement can be made either without including, or with, any resistive load. (The former gives a limiting value of gain.) Thus a sensitive Schmitt Trigger Circuit (one with low hysteresis) will show a poor rise time for a given drive. The above holds for both the clockwise hysteresis of a class-1 negative resistance in series with a resistance greater than the magnitude of $-R$, and for the counterclockwise type of class-2 and smaller series resistance[1]; i.e., in general, when the load line and the $-R$ almost coincide, transition will be slow.

Figure 8 was taken with a relatively sluggish circuit (Fig. 9) having small plate resistors; this results in low regeneration though the band width is high. This made the intermediate unstable level fall unusually high above the normal level. Also, the resulting low maximum loop gain made the initial loop gain, as a function of the applied trigger-pulse height, a more slowly varying function in the vicinity of the critical loop gain of one; i.e., the potential hill seems to fall off slowly on each side, or the $-R$ is flat. This makes for

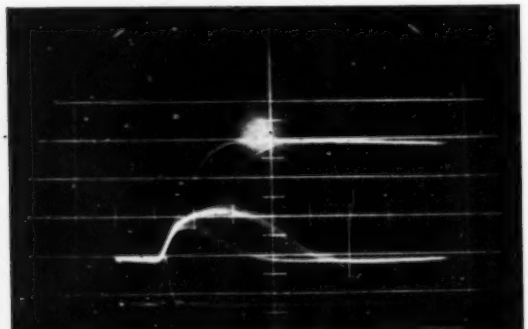


FIG. 8. Plate wave form of a multivibrator being triggered by different size pulses, all in the vicinity of one just barely adequate. The length of time the circuit remains in the intermediate state is variable just as is the length of time one can balance a cone on its point. The loop gain is close to unity at the end of the applied pulse. The intermediate state corresponds to the middle unstable $-R$ intersection, the peak of the potential hill, an infinite gain amplifier, and a loop gain of 1.00. The image of the grid pulse is superimposed. Calibration: 0.15 μsec/division, 6 volts/division on trigger and 20 volts/division elsewhere. The oscilloscope rise time was half the input pulse width.

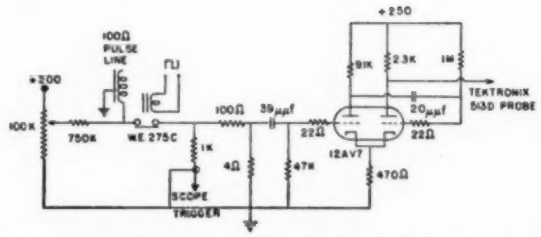


FIG. 9. Circuit displaying wave form of Fig. 8. The rectangular applied pulse is approximately 5 volts high and lasts 0.05 microsecond which is much less than the duration of the transition to the abnormal state.

greater ease in observing some action other than abrupt tripping one way or the other as the pulse height is changed slightly.

One can measure a negative resistance characteristic directly, but one is used to thinking in terms of things rolling on hills and so it is interesting to consider what is involved in drawing an equivalent potential barrier in a given case. It is obvious that the crest of the hill will correspond to the central unstable $-R$ intersection, but the negative resistance curve cannot be accurately described simply as, e.g., the derivative of a corresponding potential hill with a moat around it because of the effect of the load. This can be included in such a representation in the form of a sideward force having an angle to the vertical depending on the load, and a magnitude depending on the applied voltage. Reactive elements are related to the mass of the particle moving on this potential hill.

COUNTERS

It should be pointed out that the above considerations have little to do with the limitations on maximum permissible pulse width in binary counters and scalers. This limitation is imposed by the finite transient "switching memory" which is essentially unrelated to the steady condition of bistability. It is this memory that allows a circuit, once it has reached an otherwise symmetrical central state of equal conduction by both halves, to cross over and proceed away from its pre-

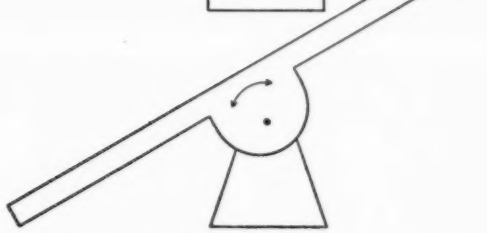


FIG. 10. A mechanical analog of a binary counter that brings out the independence of the bistable and the switching memory properties. In this case inertia gives the proper transient action, while the steady-state action is determined by the pivot location. The maximum acceptable pulse is obviously limited.

vious state. The history of counters, starting with Eccles-Jordan who did not have a counter, shows the evolution of this concept.[3–7] It is this factor of memory that accounts for the scarcity of, e.g., single-pentode counters.

The mechanical analog of a binary counter, shown in Fig. 10, may make this clear. If the center of gravity of the see-saw is above the pivot, it will be bistable. If the plunger is pushed down for enough time with enough force, but for less than the length of time required for the cross piece to become horizontal, then the cross piece will go to the other state, its inertia carrying it through the symmetrical center state. Too long a pulse will cause the switching memory to forget where it came from and make the second state unrelated to the first. In the electrical case, diode coupling or the use of double gates at the input does not in itself change this situation, as they also reach a symmetrical state of equal conduction during the switching process. The memory is still in the unequal charges stored in the "speeding-up condensers."

Thanks are to be expressed to Mr. Carter Collins for working out the detailed design of the circuit in Fig. 4 and to Mr. Harold Morris for his help in photographing the multivibrator wave shapes.

[3] W. H. Eccles and F. W. Jordan, Radio Rev. **1**, No. 3, 143 (1919).
[4] C. E. Wynn-Williams, Proc. Roy. Soc. (London), **A136**, 312 (1932).
[5] W. B. Lewis, Proc. Cambridge Phil. Soc. **33**, No. 7, 549 (1937).
[6] E. C. Stevenson, and I. A. Getting, Phys. Rev. **51**, 1027 (1937).
[7] H. Lifschutz, and J. L. Lawson, Rev. Sci. Instr. **9**, 83 (1938).

JOURNAL OF APPLIED PHYSICS VOLUME 25, NUMBER 4 APRIL, 1954

Dynamics of Corona Discharge between Cylindrical Electrodes

L. Colli, U. Facchini, E. Gatti, and A. Persano
Laboratori CISE, Via Procaccini 1, *Milano, Italy*
(Received August 17, 1953)

The qualitative explanation by Loeb of the constancy of the amplification factor with increasing corona currents with coaxial cylindrical electrodes used, as in conventional Geiger counters, is placed on a quantitative basis. As anticipated, it is shown that it is the screening action of the space charge which is responsible for the constancy and the stability of the average discharge current. The fluctuations superposed on the constant current, previously reported by us, are explained as a response of the system having a definite resonance frequency to the statistical fluctuations of the photoelectric current. The theoretical predictions are confirmed by experimental data.

1. INTRODUCTION

In previous articles[1,2] the mechanism of discharge in cylindrical counters filled with pure Argon or a mixture of Argon and CO_2 at a concentration of CO_2 up to 10^{-3} was described.

This mechanism essentially consists of the emission of ultraviolet photons in the Townsend avalanche, with the liberation of photoelectrons from the cathode, which in their turn produce a new avalanche near the wire, and so on.

The number of photoelectrons extracted at the cathode per each electron existing in the avalanche will be indicated in this paper as γ_{ph}.

The ratio between two successive avalanches is $N\gamma_{ph}$, where N is the Townsend multiplication coefficient.

If the voltage is increased, and hence the electrical field at the wire increases, N also increases, until the threshold of the corona current is reached which is characterized by the condition $N\gamma_{ph}=1$, or in other words, by the reproduction of equal successive avalanches.

The time for the reproduction of the avalanche is composed of two time intervals: (a) the time T_f corresponding to the formation and propagation of photons, and (b) the sharply defined time T_e necessary for the electrons to travel across the counter.

A study of these time intervals is presented in earlier papers in which the time T_f was found negligible compared with T_e in the case where the counters are filled by a mixture with a CO_2 concentration greater than 5×10^{-4}.[2]

Under these conditions the reproduction time is well determined, so that we will limit our analysis to this case.

Under these conditions the following factors are established: (a) The existence of a steady corona current i which increases as the overvoltage is increased. (b) The constancy of the multiplication coefficient independent of variations of the applied voltage. In the corona region N is equal to the threshold value $N_s=1/\gamma_{ph}$. (c) The presence of an oscillating component which is nearly periodic, superposed on the steady current i_0.

[1] Colli, Facchini, and Gatti, Rev. Sci. Instr. **23**, 621 (1952).
[2] L. Colli and U. Facchini, Phys. Rev. **88**, 987 (1952).

Recently Loeb has qualitatively indicated the process by which N is made constant and equal to about N_s for increasing values of the current i_0.[3]

It is the purpose of this paper to take up this analysis in order to define quantitatively the results of Loeb, and to explain the observed fluctuations of the current.

2. THE ELECTRIC FIELD IN THE COUNTER

In a counter in which there is no space charge the multiplication coefficient N is determined by the value and the shape of the electric field near the wire, and can be expressed with good approximation by

$$N=Ce^{\alpha V}, \tag{1}$$

where C and α are two characteristic constants of the counter and V is the applied voltage. The latter is related to the electric field on the wire, by

$$V=bE_b\ln(a/b), \tag{2}$$

where a is the cathode radius and b the radius of the wire.

When the counter is in the corona region, it is filled with a space charge.

We shall now study the shape of the electric field under these conditions.

The space charge is mostly made up of positive ions. In fact the current that crosses a cylinder surrounding the wire at a distance greater than the avalanche zone is composed of (1) the flow of the electrons generated at the cathode by photoelectric action and traveling towards the wire, and (2) the flow of the ions generated by electrons near the wire through the Townsend multiplication, and traveling toward the walls.

Since in our case the counter works at values of N of the order of a few hundreds, and in view of the low mobility of the positive ions, the contribution of the electrons to the space charge is almost negligible.

Let us now consider a cylindrical counter crossed by a constant current i of ions generated near the wire. Let the electric field on the wire have the value E_b'. The electric field at the distance r from the wire, E_r' as a function of the current i and of the field E_b', can be expressed by

$$E_r'=\frac{1}{r}\left[\frac{i}{2\pi\epsilon_0 K}(r^2-b^2)+b^2E_b'^2\right]^{\frac{1}{2}}, \tag{3}$$

where K, the mobility of the ions, is assumed to be constant.

According to (3), it may be noted that, in the case of operating counters, the first term under the square root sign is negligible in a zone round the wire which encloses fully the multiplication region, since this term introduces a correction of less than 0.1 percent.

The shape of E_r' in this zone is hence almost identical to the shape of the field E_r when there is no space charge.

[3] L. B. Loeb, Phys. Rev. **90**, 144 (1953).

The field E_r' in the corona region can therefore be expressed through the value of E_b' only, or, more conveniently, through a fictitious voltage V' related to E_b' by Eq. (2). Under these conditions the multiplication coefficient in the corona region can be written following Eq. (1) as

$$N'=Ce^{\alpha V'}. \tag{1a}$$

At the corona threshold the space charge is negligible, $V'=V=V_s$, and $N=N'=N_s=1/\gamma_{ph}$. γ_{ph} can be considered as roughly independent of the voltage.

We can therefore determine the constant C of relations (1) and (1a), and write that, in the corona region,

$$N'\gamma_{ph}=e^{\alpha(V'-V_s)}. \tag{4}$$

3. DYNAMICS OF THE DISCHARGE

The flow of the ions generated near the wire at the time $t-\tau$ constitutes the current $i(t-\tau)$. The corresponding charge, started from the wire in the time interval $d\tau$, is found after a time τ to be distributed close to a cylinder of radius $r=r(\tau)$. The contribution of this charge to the potential at the ends of the counter V is expressed by

$$\frac{1}{4\pi\epsilon_0}\ln\frac{a^2}{r^2(\tau)}\cdot i(t-\tau)d\tau. \tag{5}$$

The total contribution of all the space charge in the counter to the potential is given by

$$\frac{1}{4\pi\epsilon_0}\int_0^{T_i} i(t-\tau)\ln\frac{a^2}{r^2(\tau)}\cdot d\tau, \tag{6}$$

where T_i is the time taken by ions to cross the counter.

Hence

$$V=V'(t)+\frac{1}{4\pi\epsilon_0}\int_0^{T_i} i(t-\tau)\ln\frac{a^2}{r^2(\tau)}d\tau, \tag{7}$$

where V is the voltage across the counter.

If the V is kept constant by the power supply, V' is determined by the values that the current assumes during a preceding time period.

Taking into account the fact that the mechanism of the self-sustaining discharge consists of the multiplication of successive avalanches of electrons separated by the time T_e, we can write

$$i(t)=N'\gamma_{ph}i(t-T_e). \tag{8}$$

By substituting the above value of $N'\gamma_{ph}$ into (4), and in accordance with the assumption, based on experiment, that always $N'\gamma_{ph}\approx 1$ in the corona region, it follows that

$$\frac{i(t)}{i(t-T_e)}=\exp[\alpha(V'-V_s)]=1+\alpha(V'-V_s). \tag{9}$$

If the rate of change of $i(t)$ is slow enough so that

$$i(t-T_e)=i(t)-T_e\frac{di}{dt}, \tag{10}$$

we shall then have

$$\frac{di}{dt}=i\frac{\alpha}{T_e}(V'-V_s). \tag{11}$$

If we eliminate V' in (11) by substituting its equivalent obtained from Eq. (7), this results

$$\frac{di}{dt}=i\frac{\alpha}{T_e}\left\{V-V_s-\frac{1}{4\pi\epsilon_0}\int_0^{T_i} i(t-\tau)\ln\frac{a^2}{r^2(\tau)}d\tau\right\}. \tag{12}$$

It will be noted that Eq. (12) has a constant solution $i=i_0$. At this condition and according to (11), V' will be equal to V_s.

As will be seen later such solution is stable, and, if the system is excited, it will undergo damped oscillations.

On the basis of this premise, we can properly calculate the motion of the ions in the electric field expressed by Eq. (3) where E_b' has the value relative to $V'=V_s$, and i has the value i_0.

We may now write $dr/dt=KE_r'$ and integrate it so as to obtain for the motion of the ions

$$r^2(\tau)=Ai_0K^2\tau^2+2KB\tau+b^2, \tag{13}$$

where

$$A=\frac{1}{2\pi\epsilon_0 K} \quad \text{and} \quad B=\frac{V_s}{\ln(a/b)}.$$

From Eq. (13) we have

$$T_i=\frac{1}{Ai_0K}[(Ai_0a^2+B^2)^{\frac{1}{2}}-B]. \tag{14}$$

By substituting in (12) the value of $r(\tau)$ furnished by (13), we get

$$\frac{di}{dt}=i\frac{\alpha}{T_e}\left\{V-V_s-\frac{1}{4\pi\epsilon_0}\times\int_0^{T_i} i(t-\tau)\ln\frac{a^2}{Ai_0K^2\tau^2+2KB\tau+b^2}d\tau\right\}. \tag{15}$$

The steady current i_0, solution of the Eq. (15), is given by

$$V=(Ai_0a^2+B^2)^{\frac{1}{2}}+\frac{B}{2}\ln\left[\frac{(Ai_0a^2+B^2)^{\frac{1}{2}}-B}{(Ai_0a^2+B^2)^{\frac{1}{2}}+B}\cdot\frac{4B^2}{Ai_0b^2}\right]. \tag{16}$$

This relationship can also be obtained without recourse to the dynamic considerations merely by integrating Eq. (3).

From (15) it can be seen that the value of i_0 is independent of parameters α and T_e.

If we calculate the motion of ions at the applied voltage V without taking into account the space charge, the quadratic term of τ in Eqs. (13) and (15) will not appear.

In this approximation T_i is given by

$$T_i=\frac{a^2}{2KV}\ln\frac{a}{b},$$

and, for $i=i_0$ we get the following relationship, that other authors have already obtained[4]

$$V-V_s=\frac{i_0}{4\pi\epsilon_0}\frac{a^2}{2KV}\ln\frac{a}{b}=\frac{i_0}{4\pi\epsilon_0}T_i.$$

Let us consider Eq. (15) in the case of small variations of the current around i_0. For this purpose we take a new function $\Psi(t)$ related to i by

$$i=i_0e^{\Psi(t)},$$

and suppose that $\Psi(t)\ll 1$ so as to introduce in Eq. (15), under the integral sign, the approximate expression for i

$$i=i_0(1+\Psi(t)).$$

This yields

$$\frac{d\Psi}{dt}=\frac{\alpha}{T_e}\frac{i_0}{4\pi\epsilon_0}\int_0^{T_i}\Psi(t-\tau)\ln\frac{Ai_0K^2\tau^2+2KB\tau+b^2}{a^2}d\tau. \tag{17}$$

The roots of the auxiliary equation joined with this linear homogeneous equation have a negative real part. The system is stable, and, if excited by a small disturbance, it comes back to the steady current i_0 with a superposed sinusoidal damped oscillation of a well determined frequency.*

The stability of the current about its value i_0 is reflected in the stability of the multiplication coefficient N' about the value N and this theoretically justifies the hypothesis $N'=N_s$ made when Eq. (9) was written.

In view of the stability of the system we have excluded oscillations of the current of unforced type. We accordingly attributed the experimentally observed oscillations to the excitation of the damped "mode" of the system, caused by an impressed current formed by

---

[4] J. S. Townsend and P. J. Edmunds, Phil. Mag. **28**, 789 (1914); H. S. Townsend, Phil. Mag. **28**, 83 (1914); H. F. Boulind, Phil. Mag. **18**, 909 (1934); L. B. Loeb, Phys. Rev. **90**, 144 (1953).

* If we substitute, for the core $\ln(Ai_0K^2\tau^2+2KB\tau+b^2/a^2)$ in Eq. (15), the function $Fe^{-\lambda\tau}$ of τ which is also decreasing, and if we extend the limits of the integral from 0 to ∞, the equation will become exactly the differential equation

$$\Psi''+\lambda\Psi'+\frac{\alpha}{T_e}\lambda(V-V_s)(e^{\Psi}-1)=0,$$

where

$$i=i_0e^{\Psi} \quad \text{and} \quad i_0=\frac{4\pi\epsilon_0\lambda}{F}(V-V_s).$$

It can immediately be seen that this equation has only solutions which are stable with regard to disturbances of any magnitude. From this it is reasonable to deduce the stability of Eq. (15) even for great oscillations.

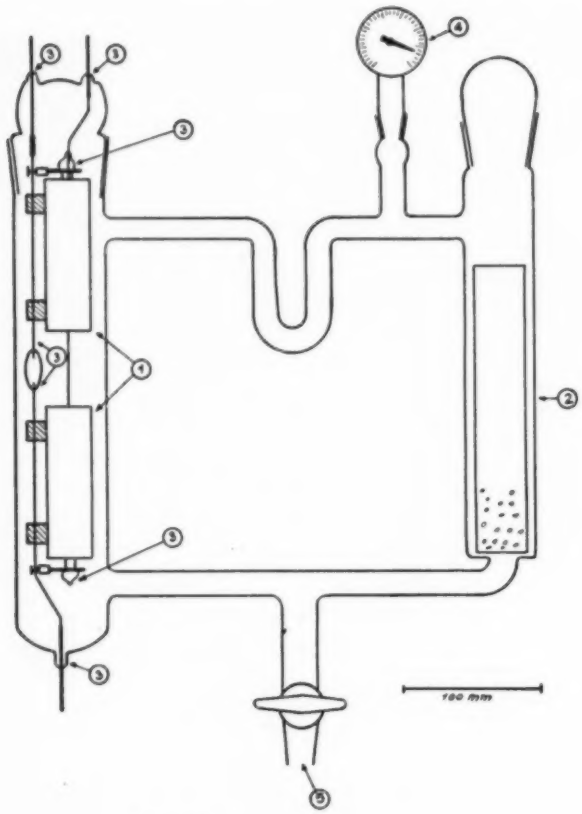


FIG. 1. Experimental assembling.

the statistical fluctuation of the photoelectron current at the cathode.

We shall therefore introduce into Eq. (9) a term which takes into account the impressed current for unit length of cylinder $I(t)/l$, where $I(t)$ is the total fluctuation current formed by electrons produced at the cathode of the counter and l is the length of the counter. Consequently, it is possible to write

$$\frac{i(t)}{i(t-T_e)}=1+\alpha(V'-V_s)+N'\frac{I(t)}{l}. \tag{18}$$

TABLE I. Experimental data.

	Set I	Set II
Cathode	Nickel, diameter 4 cm	Brass, diameter 4 cm
Anode	Tungsten, diameter 0.2 mm	Tungsten, diameter 0.2 mm
Filling	Argon+$0.7\cdot10^{-3}$ CO_2	Argon+$1.4\cdot10^{-3}$ CO_2
Total pressure	690 mm Hg	500 mm Hg
T_e	2.2 μs	1.6 μs
α	$5.12\cdot10^{-2}$ volts^{-1}	$6.8\cdot10^{-2}$ volts^{-1}
N_s	540	2500
V_s	2600 volts	2400 volts
$K(p=760$ mm Hg)	1.73 cm^2 sec^{-1} volts^{-1}	1.57 cm^2 sec^{-1} volts^{-1}

Proceeding in a similar way to that used in obtaining Eq. (15) and Eq. (17), we obtain

$$\frac{d\Psi}{dt}=\frac{\alpha}{T_e}\frac{i_0}{4\pi\epsilon_0}\int_0^{T_i}\Psi(t-\tau)\ln\frac{Ai_0K^2\tau^2+2KB\tau+b^2}{a^2}d\tau+\frac{N_s}{T_e}\frac{I(t)}{i_0l} \tag{19}$$

having assumed that $N'=N_s$ and having neglected Ψ in comparison to unity in the new term introduced.

Let us put

$$I(t)=\text{R.P. of } I_0e^{j\omega t}$$

in Eq. (19) and try to obtain a solution of the type

$$\Psi(t)=\text{R.P. of } \Psi_0e^{j\omega t}.$$

In that way we obtain for $i_0\Psi_0l/I_0$ the relation

$$\frac{li_0\Psi_0}{I_0}=\frac{j\omega(N_s/T_e)}{\left[-\omega^2-\frac{\alpha}{T_e}\frac{i_0}{4\pi\epsilon_0}G(\omega)\right]-j\left[\frac{\alpha}{T_e}\frac{i_0}{4\pi\epsilon_0}H(\omega)\right]}, \tag{20}$$

where

$$\begin{aligned}G(\omega)=\ln\frac{b^2}{a^2}&+\cos\omega x_1[Ci\omega(T_i-x_1)-Ci(-\omega x_1)]\\&+\sin\omega x_1[Si(-\omega x_1)-Si\omega(T_i-x_1)]\\&+\cos\omega x_2[Ci\omega(T_i-x_2)-Ci(-\omega x_2)]\\&+\sin\omega x_2[Si(-\omega x_2)-Si\omega(T_i-x_2)]\end{aligned} \tag{21}$$

and

$$\begin{aligned}H(\omega)=&\cos\omega x_1[Si(-\omega x_1)-Si\omega(T_i-x_1)]\\&-\sin\omega x_1[Ci\omega(T_i-x_1)-Ci(-\omega x_1)]\\&+\cos\omega x_2[Si(-\omega x_2)-Si\omega(T_i-x_2)]\\&-\sin\omega x_2[Ci\omega(T_i-x_2)-Ci(-\omega x_2)],\end{aligned} \tag{22}$$

x_1 being $=b^2/2KB$ and $x_2=-2B/Ai_0K$. $G(\omega)$ is an even function and $H(\omega)$ an odd function of ω.

The square of the absolute value of Ψ_0i_0l/I_0 can be expressed to a very good approximation as

$$\left|\frac{\Psi_0i_0l}{I_0}\right|^2=\frac{N^2}{T_e^2}\frac{\omega^2}{(\omega^2-\omega_0^2)^2+(\omega_0^4/Q^2)}, \tag{23}$$

where ω_0 is a solution of the equation in ω

$$\omega^2=-\frac{\alpha}{T_e}\frac{i_0}{4\pi\epsilon_0}G(\omega) \tag{24}$$

and

$$Q=\frac{G(\omega_0)}{H(\omega_0)}. \tag{25}$$

Equation (23) is the typical equation of a resonant system with a figure of merit Q and a resonant angular frequency ω_0.

From (21) and (22), $G(\omega)$ and $H(\omega)$ can be approximately expressed as follows:

$$G(\omega)=-\left(0.577+\ln\omega+\ln\frac{a^2}{2KB}\right), \tag{26}$$

$$H(\omega)=-1.57. \tag{27}$$

Let us now assume that the impressed current $I(t)$ is a noise current due to the fluctuations of the average current of photoelectrons $i_0 l/N_s$. The current will have the well-known spectral distribution

$$dI^2/d\omega=ei_0 l/\pi N_s, \tag{28}$$

where e is the electron charge.

The mean square value of the fluctuation of the corona current in the counter will be

$$\bar{\sigma}^2=\int_0^\infty \frac{ei_0 l}{\pi N_s}\frac{N_s^2}{T_e^2}\frac{\omega^2}{(\omega^2-\omega_0^2)^2+(\omega_0^4/Q^2)}d\omega$$

$$=\frac{ei_0 l N_s}{2\sqrt{2}T_e^2\omega_0}(1+Q^2)^{\frac{1}{2}}\left[1+\frac{1}{(1+1/Q^2)^{\frac{1}{2}}}\right]^{\frac{1}{2}}. \tag{29}$$

For the values of Q actually found (between 3 and 4), the fluctuation current will be dominantly formed by components of an angular frequency approaching ω_0 and will take a roughly sinusoidal shape of the type observed experimentally. The rapid vanishing of the expression to be integrated makes a cutting off of the integral at the higher values of ω superfluous.

Where the attenuation is not as rapid, the cutting must be introduced in order to take into account the band pass of the amplifier used in the measurements as well as the fact that, for angular frequencies much greater than ω_0, the assumption by which it was possible to write Eq. (10) no longer holds.

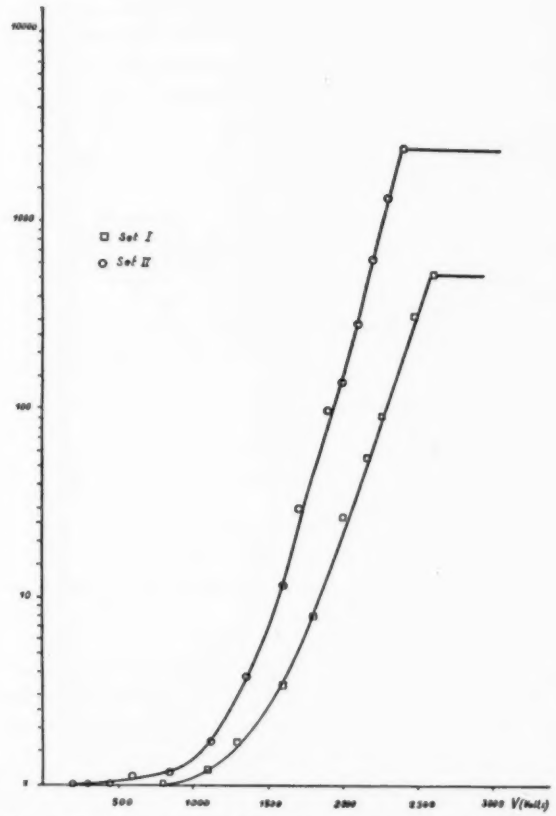


FIG. 2. Experimental multiplication curves for α particles. (Characteristics of counters as in Table I).

That the fluctuations are noticeable stems, as has been seen in the development of the calculations, from the fact that the fluctuations have been calculated from the photoelectric current emitted at the cathode, which is N_s times smaller than that flowing in the counter, and is thus open to considerable fluctuations.

Equation (20) allows us to analyze the response of the counter to the instantaneous liberation of q charges,

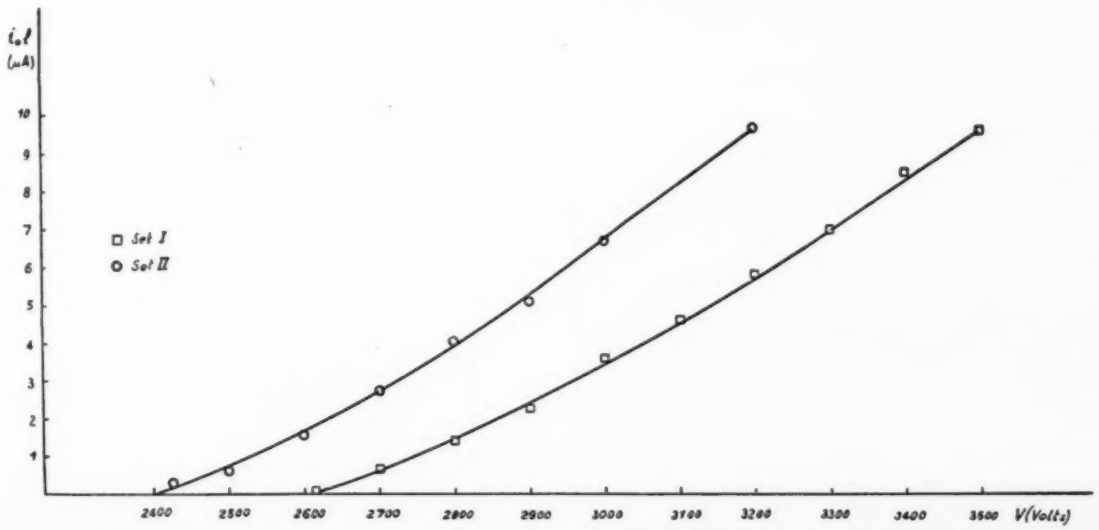


FIG. 3. Average current $i_0 l$ *versus* applied voltage (characteristics of counters as in Table I).

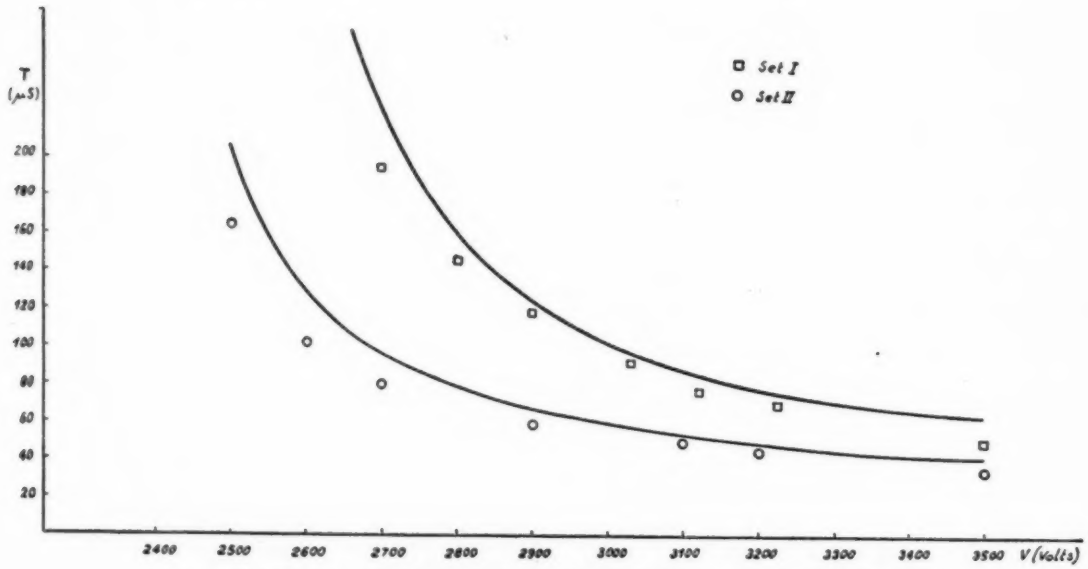


FIG. 4. Calculated and observed period of the oscillations *versus* the applied voltage. (Characteristics of counters as in Table I).

caused by an ionizing particle which has crossed it. It is enough to apply the Laplace transformation method, and take the inverse transform of the second term of (20) in order to obtain the value of $i_0\Psi(t)$ produced by the impressed current $I(t)=q\delta(t)$.

Thus (20) approximates to

$$i_0\sigma(p)=q\frac{N_s}{\omega_0 T_s}\frac{p/\omega_0}{1+(p/\omega_0 Q)+(p^2/\omega_0^2)}. \tag{30}$$

Taking the inverse transform, we obtain

$$i_0\sigma(t)=q\frac{N_s}{T_e}e^{-\omega_0 t/2Q}\left[\left(1-\frac{1}{4Q^2}\right)^{\frac{1}{2}}\cos\omega_0\left(1-\frac{1}{4Q^2}\right)^{\frac{1}{2}}t - \frac{1}{2Q}\sin\omega_0\left(1-\frac{1}{4Q^2}\right)^{\frac{1}{2}}t\right], \tag{31}$$

which represents a damped sinusoidal oscillation.

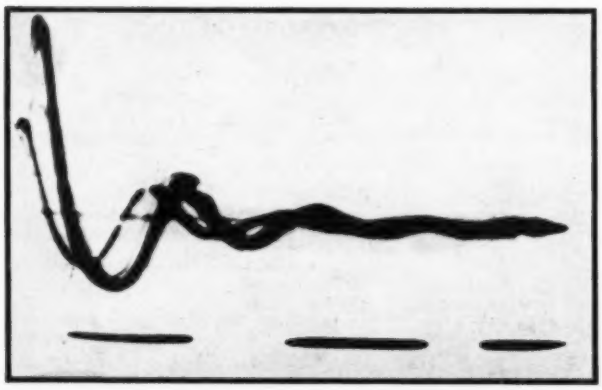


FIG. 5. α-particle pulse in corona region. Time markers 100 μsec. (Counter of set I in Table I).

The procedure above might be open to question as we have introduced the very high frequencies existing in the spectrum of $\delta(t)$. *A posteriori*, however, we can accept the solution given because it does not imply variations of a high percentage in time intervals as short as T_e, except for the discontinuity at the time 0. The discontinuity value at time 0 is however what we can expect if we take into account the fact that the charges caused by the first multiplication are just qN_s, and that they determine the average current qN_s/T_e in the time T_e before the arrival of the second avalanche.

4. EXPERIMENTAL MEASUREMENTS

The experimental check has required the careful measurement of some characteristics of the counters. For this purpose the apparatus represented in Fig. 1 was constructed. The gas mixture was purified in the

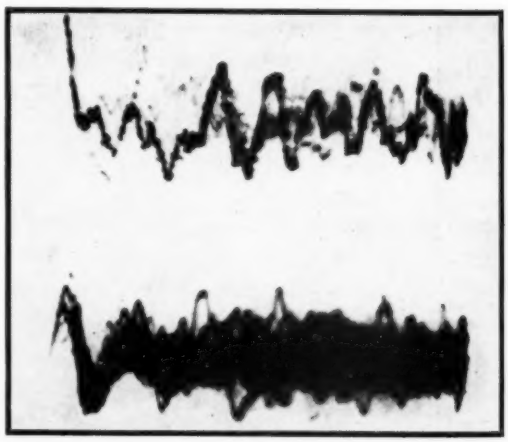


FIG. 6. Upper diagram: shape of the oscillating corona current component. Lower diagram: output voltage of the resonant circuit. Sweep time 1000 μsec.

furnace. Two sets of measurements were made. The data concerning the counters used are listed in Table I. The transit time T_e of the electrons was deduced from the time interval between successive avalanches started by an α particle, and was found to be constant over the whole range of the voltage. In Fig. 2 the experimental multiplication curves from which the values of α were deduced, are plotted. In Fig. 3, the experimental values of the average current i_0 *versus* the applied voltage V are shown as points, and compared with the theoretical curves. In calculating these curves, using Eq. (16), the mobility value was so adjusted as to fit the highest experimental point reached. The values of the mobility reduced to normal pressure, are shown in Table I, and appear compatible with those obtained directly by other authors.

In Fig. 4 the experimental values of the period of the oscillating components are indicated as points and compared with the theoretical curves of the resonant period $2\pi/\omega_0$, calculated according to Eq. (24).

In Table II are shown the experimental values and theoretical values of the amplitude of the oscillating current, calculated according to Eq. (29).

The experimental values of the period and amplitude have been obtained by observing on the oscillograph the voltage drop across an impedance formed by a resistance shunted with capacitance connected in series with the counter.

The measurements of σ are not too reliable, but indicate a satisfactory order of magnitude agreement. The calculated values of Q were always between 3 and 4 in both sets of measurements.

In Fig. 5 is shown the oscillogram of the current excited by a single α particle. The corresponding theoretical equation is (31). The experimental values of Q, of which no precise measurement was possible, because the fact that the disturbances produced by the α particles were not "small," agree within 30 percent with the theoretical values.

In Fig. 6 are put together for comparison an oscillogram of the current in the counter and that of the voltage at the ends of a parallel oscillating circuit, with $Q=3.5$, excited by a current generator having a uniform spectrum along 4 octaves above and below the resonant frequency. The analogy in the shapes of the two oscillograms is clear.

TABLE II. Comparison between theoretical and experimental data about the amplitude of the fluctuation current $\bar{\sigma}$.

	Set I		Set II	
Voltage	Experimental	Theoretical	Experimental	Theoretical
2775	$4.25\cdot10^{-8}$ A	$3.0\cdot10^{-8}$ A		
3000	$5.4\cdot10^{-8}$ A	$4.25\cdot10^{-8}$ A		
3500	$4.6\cdot10^{-8}$ A	$5.3\cdot10^{-8}$ A		
2600			$9.65\cdot10^{-8}$ A	$8.4\cdot10^{-8}$ A
2900			$11.0\cdot10^{-8}$ A	$12.2\cdot10^{-8}$ A
3200			$12.6\cdot10^{-8}$ A	$15.1\cdot10^{-8}$ A

ACKNOWLEDGMENTS

We wish to thank Professor G. Bolla for his kind interest in the present work and Dr. A. Ascari for valuable suggestions.

We thank also Professor L. B. Loeb for his constant interest and stimulating advice.

Focusing of an Electron Beam by Periodic Fields

A. M. CLOGSTON AND H. HEFFNER
Bell Telephone Laboratories, Murray Hill, New Jersey
(Received July 27, 1953)

The problem of focusing long electron beams by means of magnetic or electric fields which vary periodically along the beam is considered. Four specific cases are discussed: periodic fields of the axially symmetric and quadrupole type with either electric or magnetic fields employed. The equations of motion are written and solutions obtained which show beam trajectories corresponding to essentially parallel flow. Under certain specified conditions only small ripples are present. Actual trajectories as plotted by the analog computer are shown and charts and equations are presented to aid in the design of practical periodic focusing systems.

INITIATED by the discovery of the "Strong Focusing Effect" at the Brookhaven National Laboratory by Courant, Livingston, and Snyder,[1] considerable attention has been given recently to the possibilities of focusing electron beams by means of periodic electric and magnetic fields. The work at Brookhaven has been concerned with confining a very rarefied stream of protons with initially divergent velocities to a narrow beam over a very long path. By the use of fields that cause the beam to alternately converge and diverge, it has been found that stable orbits for the protons can be achieved that lie within a much smaller beam than possible with previous methods. With electron beams one is most interested in cases where the defocusing effect of the space charge is the important consideration. The presence of space charge completely changes the character of the orbits, but we find nevertheless that such an electron stream can also be confined into a narrow beam by the use of periodically varying fields. Four particular instances of periodic focusing have turned out to be most interesting and will be discussed herewith. They are the cases of (a) an axially symmetric periodic magnetic field, (b) a quadrupolar periodic magnetic field, (c) an axially symmetric periodic electric field, and (d) a quadrupolar periodic electric field.

One may well ask why focusing by periodic fields is of interest particularly since it will turn out that in case (a) for example, the peak magnetic field required is on the order of one and a half times that necessary for nonperiodic focusing. The answer is that periodic focusing may be useful for one or more of at least three reasons: (1) the reduction of magnet weight effected by periodic magnetic focusing when permanent magnets are employed,[2] (2) the possibility of making the focusing electrodes an integral part of the rf structure in certain traveling wave tubes, and (3) the production of an electron beam with a periodically varying boundary, charge density, or velocity.

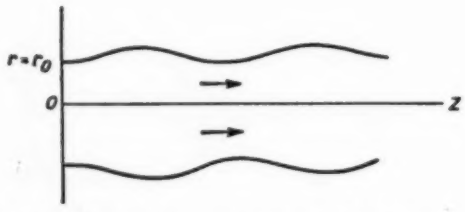


FIG. 1. Cylindrical electron beam in axially symmetric magnetic field.

[1] Courant, Livingston, and Snyder, Phys. Rev. 88, 1190–1196 (1952).

A. AXIALLY SYMMETRIC PERIODIC MAGNETIC FIELD

Let us consider an electron stream proceeding along the z axis as shown in Fig. 1. We will suppose that an axially symmetric magnetic field is present with a z component

$$B=B_0\cos kz, \tag{1}$$

where $k=2\pi/p$ and p is the period of the magnetic field. This form of field variation could be approximated along the axis of a series of oppositely poled ring magnets which had been magnetized in the z direction. If we suppose that the electrons have emerged from a source outside the magnetic field, the angular velocity of the electrons (measured anti-clockwise looking along the positive z direction) is $\dot{\theta}=-\eta B/2$. The radial equation of motion becomes

$$\ddot{r}+r\left(\frac{\eta B}{2}\right)^2+\eta E_r=0, \tag{2}$$

where E_r is the radial electric field and η is the charge to mass ratio of the electron.

The motion of the electrons in this periodic magnetic field is a complicated one, and to progress we must make some drastic assumptions. Thus, we will assume that, (1) The electrons on the outside surface of the beam always remain the outermost electrons. (2) Axial variations of the beam occur in a distance large compared to the diameter of the beam. (3) Axial

[2] This weight saving occurs in spite of the larger peak field required in the axially symmetric case. It may be seen simply by consideration of the work done in assembling a large number of magnets a great distance away first into a uniform field and second into a periodic north-south field. In the first case work is done on the mover while in the second case the mover must do work. The energy thus represented is in the first case abstracted from and in the second case added to the magnetic field, indicating that for a given weight of magnet the field produced will be larger if the magnets are of opposing polarity. In actuality, the ratios of maximum fields produced for the same magnet weight may be a factor of five or ten.

velocities of the electrons are constant throughout the beam. Using the second assumption, we can write for the radial electric field on the outside surface of the beam

$$2\pi r\epsilon_0 E_r = \int \rho ds, \tag{3}$$

where ϵ_0 is the dielectric constant of free space, ρ is the charge density, and ds is the element of area. Because of assumption (1) and (3), $\int \rho ds$ must be a constant. Let us then suppose that the radius of the beam at $z=0$ is r_0 and define an average charge density at this point by

$$\int \rho ds = \pi r_0^2 \rho_{Av}. \tag{4}$$

Let us also define the plasma frequency

$$\omega_p^2 = -\eta\rho_{Av}/\epsilon_0. \tag{5}$$

We then have for E_r

$$E_r = -\frac{1}{2\eta}\frac{r_0^2\omega_p^2}{r}. \tag{6}$$

The equation of motion for the outermost electrons becomes, consequently,

$$\ddot{r} + r\left(\frac{\eta B}{2}\right)^2 - \frac{r_0^2\omega_p^2}{2r} = 0. \tag{7}$$

Let us set $\langle\omega_c^2\rangle_{Av} = \frac{1}{2}(\eta B_0)^2$ where $B_0/\sqrt{2}$ is the rms magnetic field.[3] Then we have

$$\frac{d^2r}{dt^2} + r\frac{\langle\omega_c^2\rangle_{Av}}{2}\cos^2 kz - \frac{r_0^2\omega_p^2}{2r} = 0. \tag{8}$$

If v is the constant velocity of the electrons, we can set $t=z/v$. Let us also place $kv=\omega$, where ω is the radian frequency at which the field appears to alternate as one moves with the electrons. In addition, we will place $kx=x$, and $kr=y$. Equation (8) then becomes

$$\frac{d^2y}{dx^2} + \left[\frac{1}{2}\frac{\langle\omega_c^2\rangle_{Av}}{\omega^2}\cos^2 x\right]y - \frac{1}{2}\left(\frac{\omega_p}{\omega}\right)^2\frac{y_0^2}{y} = 0. \tag{9}$$

We now place

$$A = -\frac{1}{2}\left(\frac{\omega_p}{\omega}\right)^2, \quad B = -\frac{1}{4}\frac{\langle\omega_c^2\rangle_{Av}}{\omega^2}, \quad \xi = y/y_0$$

and expand the cosine term to obtain

$$\frac{d^2\xi}{dx^2} + B(1+\cos 2x)\xi - \frac{A}{\xi} = 0. \tag{10}$$

[3] If B does not vary with distance, we should correspondingly set $\langle\omega_c^2\rangle_{Av} = (\eta B)^2$. We see then from Eq. (7) that a solution exists where $r=r_0$ if $\langle\omega_c^2\rangle_{Av} = 2\omega_p^2$. This type of motion is known as Brillouin Flow.

We have chosen the magnetic field to be a maximum at $x=0$. Let us introduce the electrons at this point with parallel motion so that $\xi_0=1$ and $(d\xi/dx)_0=0$.

We can obtain an approximate solution of Eq. (10) for a special relation between A and B by assuming that ξ is nearly unity. Let us place $\xi=1+\sigma$ where σ is small. Making this assumption in Eq. (10) and linearizing the term in $1/\xi$ we obtain

$$\frac{d^2\sigma}{dx^2} + [(B+A)+B\cos 2x]\sigma + [(B-A)+B\cos 2x] = 0. \tag{11}$$

Let us now place $\sigma=\delta(1-\cos 2x)+\epsilon$. If we set

$$\delta = B/(A-4) \tag{12}$$

and

$$\delta = \frac{A-B}{A+(B/2)}, \tag{13}$$

we can show by substitution in Eq. (11) that ϵ is of order δ^2. Inspection of Eqs. (12) and (13) shows that A must be a small quantity and nearly equal to B if δ is to be small. Making this assumption we find

$$B = A[1+\tfrac{3}{8}A] \tag{14}$$

so that

$$\langle\omega_c^2\rangle_{Av} = 2\omega_p^2\left[1+\frac{3}{16}\left(\frac{\omega_p}{\omega}\right)^2\right] \tag{15}$$

$$\delta = -\frac{1}{8}\left(\frac{\omega_p}{\omega}\right)^2. \tag{16}$$

The final equation for the orbit of the outermost electrons can then be written

$$r = r_0\left[1-\frac{1}{8}\left(\frac{\omega_p}{\omega}\right)^2(1-\cos 2kz)\right], \tag{17}$$

provided that $\langle\omega_c^2\rangle_{Av}$ satisfies Eq. (15) and $\frac{3}{16}(\omega_p/\omega)^2 \ll 1$. We note also that this solution exhibiting nearly parallel flow of the electrons requires that the beam enter at a point of maximum field with no transverse motion. This is not a difficult condition to achieve experimentally.

We now must discuss the behavior of the stream when B does not satisfy Eq. (14). For small departures from Eq. (14) we can obtain a rough idea of how things go by ignoring the terms in $\cos 2x$ in Eq. (11). The equation for σ then becomes

$$\frac{d^2\sigma}{dx^2} + (B+A)\sigma + (B-A) = 0. \tag{18}$$

The appropriate solution of this equation is

$$\sigma = \frac{A-B}{A+B}[1-\cos(A+B)^{\frac{1}{2}}x]. \tag{19}$$

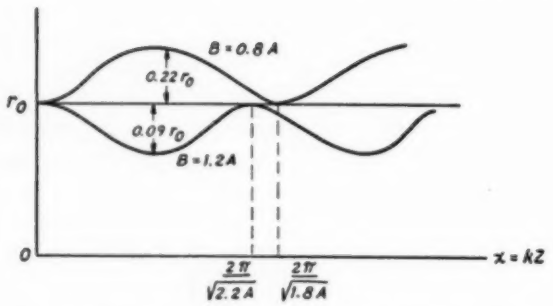


FIG. 2. Beam scallops away from the balance condition.

When $A=B$, to this rough approximation, the orbit is a straight line at radius $r=r_0$. When B is slightly different from A the orbits are as shown in Fig. (2).[4]

We see that deep scallops tend to develop in the beam for B not equal to A. As B gets progressively smaller the period of the scallops increases until eventually the orbit approaches the usual spreading curve determined by Eq. (10) with $B=0$.

As B becomes progressively larger a very interesting situation develops. We see that the scallops tend to become deeper and it is clear that the radial velocities will become increasingly great. In this case, a time will come when we can ignore the space charge term in Eq. (10) and set $A=0$. We have then for $B\gg A$,

$$\frac{d^2\xi}{dx^2}+B(1+\cos 2x)\xi=0. \tag{20}$$

Equation (20) is a specialized Mathieu Equation, and the motion it describes is well known.[5] In particular, it is characterized by stable and unstable zones depending upon the parameter B as shown in Fig. 3, where the unstable zones are shaded.

Usually A will be small enough to lie well within the first stable zone. The general features of the motion as B, the magnetic field parameter, increases from zero can now be described as follows:

(1) For B zero, the electrons maintain laminar flow and the beam spreads in the usual way.

(2) For B small, the beam will spread out, but after a long excursion return to its original diameter if unhindered.

(3) As B increases, this outward motion will become of smaller period and amplitude until nearly parallel motion is obtained for $B=A$.

(4) As B becomes larger than A, deep inward scallops develop in the beam until eventually the inward velocities become large, space charge becomes unimportant, and laminar flow is destroyed. At this point, the orbits of the electrons cross the axis but are not lost from the beam although the excursions may become large.

(5) When B becomes approximately 0.7, the flow becomes unstable and the electrons will be rapidly lost from the beam.

(6) As B becomes still larger further stable and unstable zones appear. In the stable zones the orbits rapidly cross and recross the axis.

The above discussion in essence has revealed the following. To obtain a stable, nearly parallel, flow we must have, broadly speaking,

$$\omega_p<\omega \tag{21}$$

and

$$\langle\omega_c^2\rangle_{\text{Av}}=2\omega_p^2. \tag{22}$$

It may be worthwhile at this point to explore the conditions under which assumption 3 (constant axial velocity) is justified. Let us suppose that the outer radius of the beam at some point z is R, while the outer radius at $z=0$ is R_0. Suppose the velocity of all electrons at z is v, but we allow v to vary along the beam. Now, the current in the beam is constant so we can write at any point

$$\int \rho ds=I/v, \tag{23}$$

and therefore by Eq. (3) we have

$$\left(\frac{\partial V}{\partial r}\right)_R=-\frac{I}{2R\epsilon_0 v}=\frac{b}{R(V_R)^{\frac{1}{2}}}, \tag{24}$$

where $b=-3.03\times10^4 I$. Now, if conditions vary slowly along the beam, we can write for the potential outside the beam

$$V=a\ln\frac{r}{r_a}+V_a \tag{25}$$

and

$$\partial V/\partial r=a/r, \tag{26}$$

where V_a is the potential on a cylindrical shield at radius r_a. We write, therefore, at $r=R$

$$\left(\frac{\partial V}{\partial r}\right)_R=\frac{a}{R}, \tag{27}$$

$$VR=a\ln\frac{R}{r_a}+V_a. \tag{28}$$

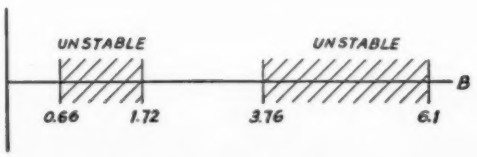


FIG. 3. Regions of instability.

[4] A solution of the exact equation on the Laboratories' computer reveals that the argument of the cosine term is not $(A+B)^{\frac{1}{2}}x$ but is rather $(2B)^{\frac{1}{2}}x$ meaning that the scallop period is $\sqrt{2}$ times the rms cyclotron period. Note that this is the same perturbation periodicity as occurs in Brillouin flow. The amplitude of the scallops is shown in Fig. 6.

[5] See for example N. W. MacLachlan, *Theory and Applications of Mathieu Functions* (Oxford University Press, New York, 1947).

From Eqs. (27) and (24), we have

$$a=\frac{b}{(V_R)^{\frac{1}{2}}} \tag{29}$$

and

$$V_R=\frac{b}{(V_R)^{\frac{1}{2}}}\ln\frac{r}{r_a}+V_a. \tag{30}$$

We can write Eq. (30) as

$$(V_R)^{\frac{1}{2}}(V_R-V_a)=b\ln\frac{R}{r_a}. \tag{31}$$

Suppose the potential at R_0 is V_0 and set $\Delta V=V_R-V_0$ and $\Delta R=R-R_0$. There

$$\frac{\Delta V}{V_0}=\frac{b}{V_0^{\frac{1}{2}}}\frac{\Delta R}{R_0} \tag{32}$$

$$=\alpha\frac{\Delta R}{R_0}, \tag{33}$$

where $\alpha=-3.03\times10^4 I V_0^{-\frac{1}{2}}$. If $I=10^{-3}$ amp and $V_0=900$ volts, $\alpha=0.0006$.

Let us now return to Eq. (8). If v varies along the beam, the equation should be written

$$\frac{d^2R}{dt^2}+R\frac{\langle\omega_c^2\rangle_{Av}}{2}\cos^2kz-\left(\frac{V_0}{V_R}\right)^{\frac{1}{2}}\frac{R_0^2\omega_p^2}{2R}=0. \tag{34}$$

If we set $\omega=kv_0$ and use Eq. (32) this can be written, analogously to Eq. (10), as

$$\frac{d^2\xi}{dx^2}+\frac{\alpha}{2}\left(\frac{d\xi}{dx}\right)^2+\left(\frac{V_0}{V_R}\right)B(1+\cos2x)\xi$$

$$-\left(\frac{V_0}{V_R}\right)^{\frac{1}{2}}\frac{A}{\xi}=0. \tag{35}$$

It is now observed that α is negligibly small and V_0 is negligibly different from V_R, so that Eq. (35) reduces to Eq. (10).

The detailed behavior of the orbital solutions of Eq. (10) has been investigated using the laboratories' differential analyzer. In general, these results justify the approximations resorted to in obtaining the explicit solutions presented above.

Before summarizing the computer results we shall indicate typical values of the space-charge parameter,

$$A=\frac{1}{2}\frac{\omega_p^2}{\omega^2}$$

and the magnetic field parameter,

$$B=\frac{1}{4}\frac{\langle\omega_c^2\rangle_{Av}}{\omega^2}.$$

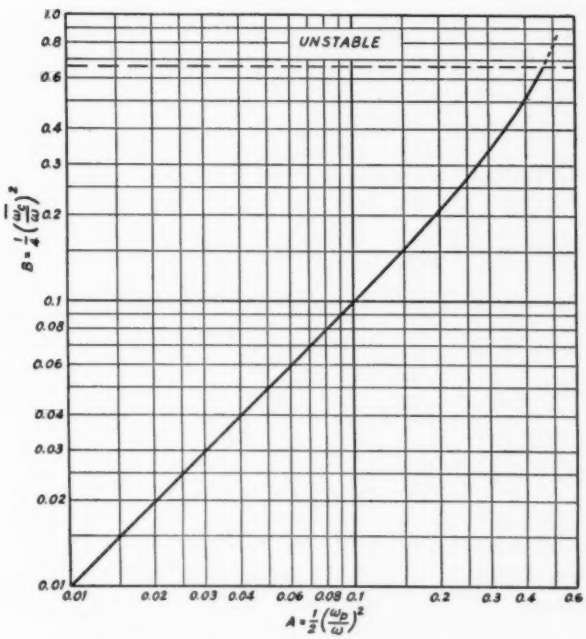


FIG. 4. Relation between the space-charge and magnetic-field parameters for minimum ripple.

Suppose we have a vacuum tube with the following characteristics:

Beam voltage	= 1000 volts
Beam current	= 1 ma
Beam diameter	= 0.080 in.
Magnet periodicity	= 1.00 in.

We then find that

$$\omega^2=2.15\times10^{19},$$

$$\omega_p^2=2.62\times10^{18},$$

and the space charge parameter

$$A=0.0762.$$

For equilibrium, then, the magnetic-field parameter by relation (14) is

$$B=0.0764,$$

indicating that the root mean square magnetic field required is 145 gauss.

It should be pointed out that the space charge parameter may be written in terms of the perveance K, the magnet periodicity p, and the initial beam diameter d as

$$A=1540K(p/d)^2.$$

The magnetic field parameter may be written as

$$B=\frac{3.60\times10^{-3}\langle B^2\rangle_{Av\,gauss}p^2_{inches}}{V_{volts}}.$$

Note that for a given diode electron gun a change in beam voltage leaves the space charge parameter A,

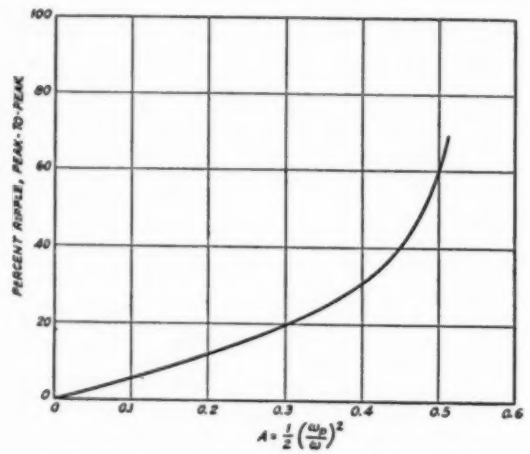


FIG. 5. Amount of ripple at the balance point.

unchanged while the parameter B does change and is in fact inversely proportional to the beam voltage. A change in the magnet periodicity p changes each parameter in the same proportion.

The following figures present a summary of the computer results. Figure 4 shows the relation between the magnetic field parameter B and the space charge parameter A required for minimum ripple. Figure 5 indicates the peak-to-peak magnitude of the ripple at this balance point. Figure 6 gives an indication of the magnitude of the ripple away from the balance point. The following figures are tracings of the actual orbits as plotted by the differential analyzer. The first group, Figs. 7(a) through 7(g) indicate the effect on the beam diameter when the plasma frequency (A parameter) is changed while the magnetic field is held constant (B parameter). Figure 8 illustrates what happens if the beam enters the periodic focusing region initially converging or diverging. The variables are such that unconfined the beam would spread to $2\pi[d\xi/dx]_0$ times the initial radius in a magnet period.

B. QUADRUPOLAR PERIODIC MAGNETIC FIELD[6]

We shall next discuss the focusing of an electron beam by a magnetic field with four-pole rather than axial symmetry. A possible magnet configuration to give this field is indicated in Fig. 9. In a plane perpendicular to the z axis these fields will appear about as shown in Fig. 10.

Let us first develop the expression we will need for the magnetic fields. We can write for the magnetic potential

$$\Phi=-\frac{B_0}{2a}\cos kz\,\sin 2\Phi r^2\left[1+\frac{(kr)^2}{12}+\cdots\right]. \tag{36}$$

As usual $k=2\pi/p$ and therefore $kr=2\pi(r/p)$. Let us suppose that the period of the field is great enough compared to the greatest radius of the beam so that we can neglect terms of order $(kr)^2$ and higher within the bracket. We can then write for the fields

$$B_r=-\frac{B_0}{a}r\cos kz\,\sin 2\Phi, \tag{37}$$

$$B_\Phi=-\frac{B_0}{a}r\cos kz\,\cos 2\Phi, \tag{38}$$

$$B_z=+\frac{kB_0}{2a}r^2\sin kz 2\Phi, \tag{39}$$

where B_0 is the magnitude of the field at $z=0$ and radius a.

In this problem we shall find it much more convenient to work in (xyz) coordinates rather than $(r\Phi z)$. Accordingly, we will rewrite the fields as follows:

$$B_x=-\frac{B_0}{a}y\cos kz, \tag{40}$$

$$B_y=-\frac{B_0}{a}x\cos kz, \tag{41}$$

$$B_z=k\frac{B_0}{a}xy\sin kz. \tag{42}$$

Let us next consider how we shall express the electric force on the electrons. We must again make the three assumptions used in Sec. A. These are, however, no longer sufficient to specify the force on the outer electrons since we no longer have a problem with axial symmetry, and we must therefore introduce some additional ideas. Consider first the force on the electrons from the magnetic field. If the components of velocity are v_x, v_y, and v_z, and the components of force are M_x, M_y, and M_z, we have

$$M_x=-e\frac{B_0}{a}[v_y ky\,\sin kz+v_z\cos kz]x, \tag{43}$$

$$M_y=e\frac{B_0}{a}[v_x kx\,\sin kz+v_z\cos kz]y. \tag{44}$$

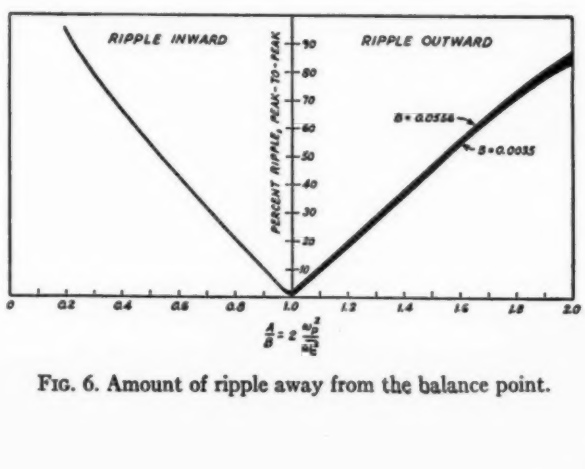


FIG. 6. Amount of ripple away from the balance point.

[6] Courant, Livingston, and Snyder (see reference 1) have described this type of focusing as it applies to synchrotron focusing for which space charge defocusing is negligible.

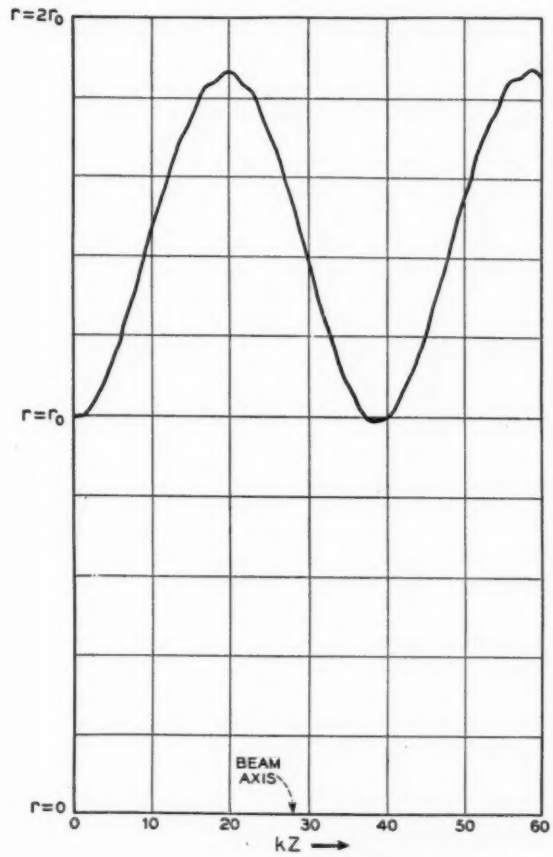


FIG. 7(a). Periodic axially symmetric magnetic focusing. Plasma frequency too high. $A=0.0278$ $B=0.0139$.

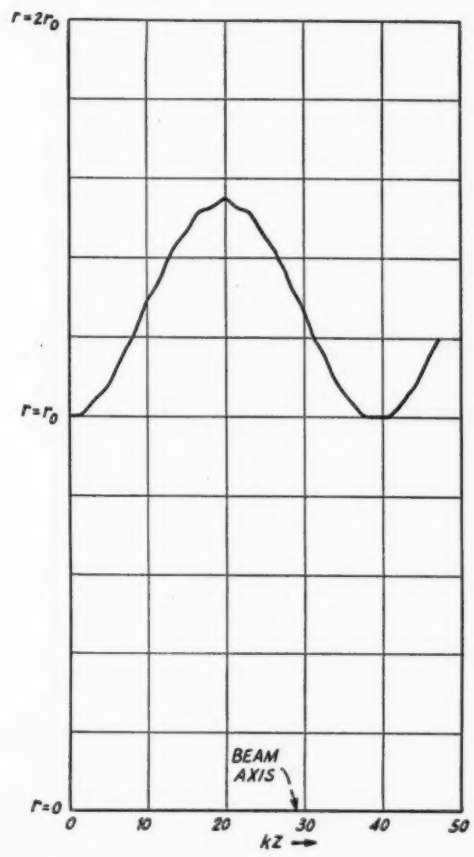


FIG. 7(b). Periodic axially symmetric magnetic focusing. Plasma frequency too high. $A=0.0222$ $B=0.0139$.

Now, by assumption (2), v_z and v_y are small compared to v_x, and for the same reason we have assumed kx and ky small compared to unity. We shall therefore write

$$M_x=-e\frac{B_0}{a}v\cos kzx, \tag{45}$$

$$M_y=e\frac{B_0}{a}v\cos kzy, \tag{46}$$

where v is used for v_x. It will now be observed that the magnetic force in the x direction is just 180° out of phase with that in the y direction. We will therefore surmise that the cross section of the beam varies along z as shown in Fig. 11. In other words, the initial magnetic force on the electrons in the x direction is negative and tends to squeeze the beam, while in the y direction the force is positive and tends to extend the beam.

We shall now confine our attention to an electron on the outer surface of the beam at $x=0$ and $y=y_0$ as shown in Fig. 12(a). We wish to find out how E_y measured at this particular electron varies as we move down the beam. Let us suppose that the charge density in the beam is originally uniform with a value ρ. Then the field at that point is given by

$$E_y=(\rho/2\epsilon)y_0. \tag{47}$$

Suppose that ρ stays constant as the cross section changes. Let the electron be in a position y as shown in Fig. 12(b). Suppose that the charge density ρ actually filled a circle of radius y. Then the field would be increased over Eq. (47) by a factor y/y_0, and the true field must clearly be less than this. Suppose, on the other hand, all the charge in the beam were distributed uniformly into a circle of radius y. Then the field would be decreased by a factor y_0/y, and the true field must clearly be greater than this. We therefore will assume with considerable accuracy that the true field does not change at all during the motion. Remembering the definition of ω_p^2, we write

$$\eta E_y=-\frac{y_0}{2}\omega_p^2. \tag{48}$$

For the y equation of motion we can now write from Eqs. (46) and (48),

$$\frac{d^2y}{dt^2}-\left[\frac{\eta B_0}{a}\cos kz\right]vy-\tfrac{1}{2}\omega_p^2y_0=0. \tag{49}$$

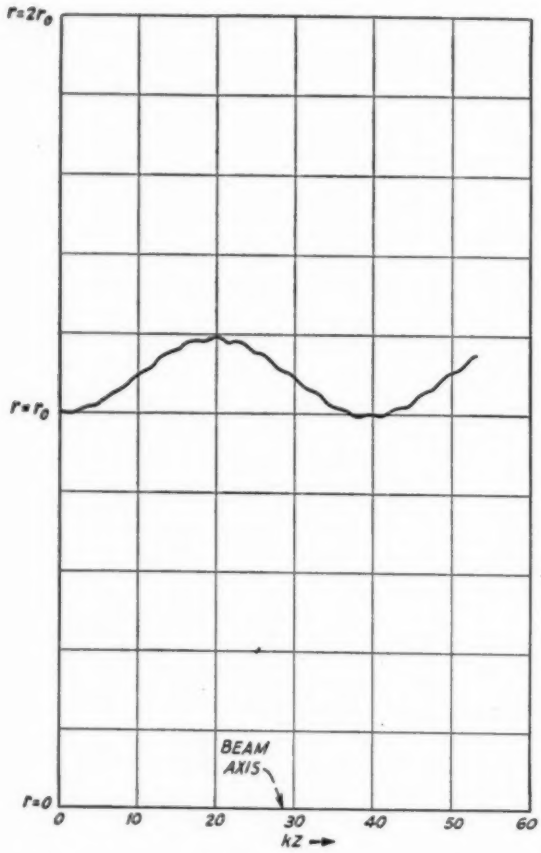


FIG. 7(c). Periodic axially symmetric magnetic focusing. Plasma frequency too high. $A=0.0166$ $B=0.0139$.

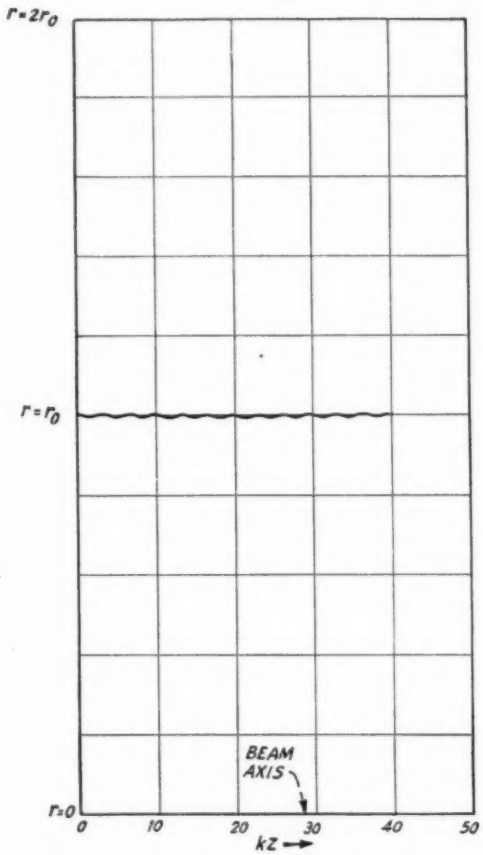


FIG. 7(d). Periodic axially symmetric magnetic focusing. Balance condition. $A=0.139$ $B=0.0139$.

Similarly, for the x equation of motion of an electron on the outer surface of the beam at $y=0$ we obtain

$$\frac{d^2x}{dt^2}+\left[\frac{\eta B_0}{a}\cos kz\right]vx-\tfrac{1}{2}\omega_p{}^2x_0=0. \tag{50}$$

Let us now set

$$\begin{aligned}\xi&=y/y_0 \quad \zeta=x/x_0\\ q&=kz\\ \omega&=kv\\ \omega_c&=\eta B_0.\end{aligned} \tag{51}$$

Then Eqs. (49) and (50), respectively, become

$$\frac{d^2\xi}{dq^2}-\left[\frac{\omega_c v}{\omega^2 a}\cos q\right]\xi-\frac{1}{2}\left(\frac{\omega_p}{\omega}\right)^2=0, \tag{52}$$

$$\frac{d^2\zeta}{dq^2}+\left[\frac{\omega_c v}{\omega^2 a}\cos q\right]\zeta-\frac{1}{2}\left(\frac{\omega_p}{\omega}\right)^2=0. \tag{53}$$

If we place

$$A=\frac{1}{2}\left(\frac{\omega_p}{\omega}\right)^2, \tag{54}$$

$$B=\frac{\omega_c v}{\omega^2 a}. \tag{55}$$

Equations (52) and (53) become

$$\frac{d^2\xi}{dq^2}-[B\cos q]\xi-A=0, \tag{56}$$

$$\frac{d^2\zeta}{dq^2}+[B\cos q]\zeta-A=0. \tag{57}$$

We will again be interested in solutions of (56) and (57) which exhibit very nearly parallel flow. Let us again assume that the beam enters parallel, that is at $q=0$, $\xi=\zeta=1$ and $d\xi/dq=d\zeta/dq=0$ and take as approximations to the orbits,

$$\xi=1+\epsilon(1-\cos q) \quad \epsilon\ll 1, \tag{58}$$

$$\zeta=1-\delta(1-\cos q) \quad \delta\ll 1. \tag{59}$$

We may substitute these into the differential equations (56) and (57), drop the higher frequency cosine terms, and obtain the necessary balance conditions. From (56) and (58) we find

$$B^2/(1-B)=2A, \tag{60}$$

$$\epsilon=2A/B, \tag{61}$$

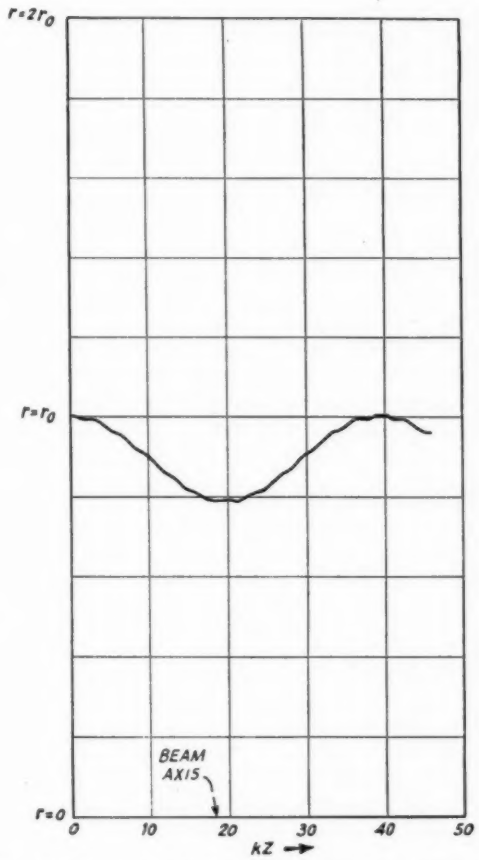


FIG. 7(e). Periodic axially symmetric magnetic focusing. Plasma frequency too low. $A=0.0111$ $B=0.0139$.

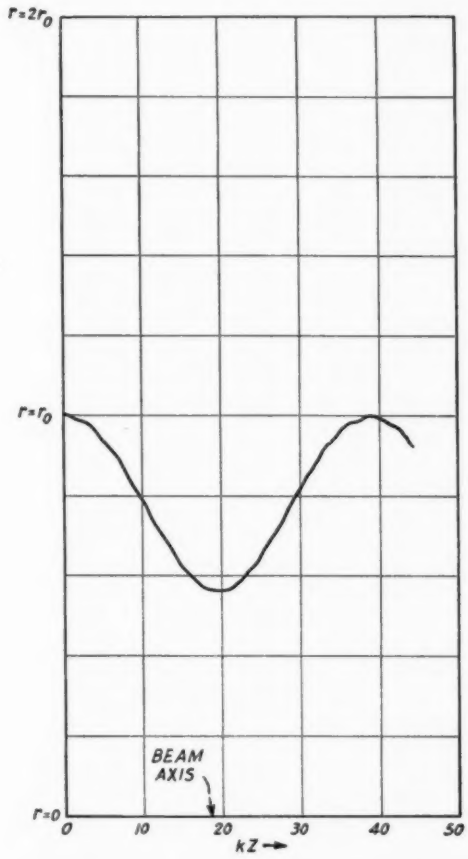


FIG. 7(f). Periodic axially symmetric magnetic focusing. Plasma frequency too low. $A=0.0038$ $B=0.0139$.

and from (57) and (59)

$$B^2/(1+B)=2A, \tag{62}$$

$$\delta=2A/B. \tag{63}$$

We observe that Eqs. (60) and (62) are incompatible unless B approaches zero. We cannot, therefore, expect to find solutions of the type proposed unless B is very small. Let us assume this to be the case. We have then that

$$B=(2A)^{\frac{1}{2}} \tag{64}$$

and

$$\epsilon=\delta=(2A)^{\frac{1}{2}}. \tag{65}$$

Substituting from Eqs. (54) and (55) we obtain

$$\omega_c v/\omega^2 a=(\omega_p/\omega), \tag{66}$$

$$\epsilon=(\omega_p/\omega) \tag{67}$$

so that the orbit of the outer electrons at the balance condition (64) is

$$y=y_0[1+(\omega_p/\omega)(1-\cos kz)]. \tag{68}$$

The implications of Eqs. (66) and (67) become clearer if we accept the plasma frequency ω_p, the beam velocity v, and the allowable beam excursion ϵ, as quantities fixed by external circumstances. The strength and periodicity of the magnetic field compatible with these restrictions are then

$$p=2\pi\epsilon v/\omega_p, \tag{69}$$

$$\omega_c/a=\omega_p^2/v\epsilon. \tag{70}$$

Let us now discuss what happens when B departs from the critical value determined by Eq. (62). For very small B, the assumptions leading to Eq. (62) are no longer tenable and the orbits should be described by Eq. (10) with $B=0$. We would have then the usual spreading of a beam caused by space charge. When B becomes considerably larger than the critical value, we may again ignore the space charge and obtain the equations

$$\begin{aligned} d^2\xi/dq^2-(B\cos q)\xi&=0,\\ d^2\zeta/dq^2+(B\cos q)\zeta&=0. \end{aligned} \tag{71}$$

These are again specialized Mathieu equations and will exhibit regions of stability and unstability as shown in Fig. 13, with the unstable regions shaded as before.

The first unstable zone appears at $B=0.44$. This corresponds to a value of ω_c given by

$$\frac{\omega_c}{a}=0.44\frac{\omega^2}{v}=\frac{0.44}{\epsilon}\left(\frac{\omega_p^2}{v\epsilon}\right). \tag{72}$$

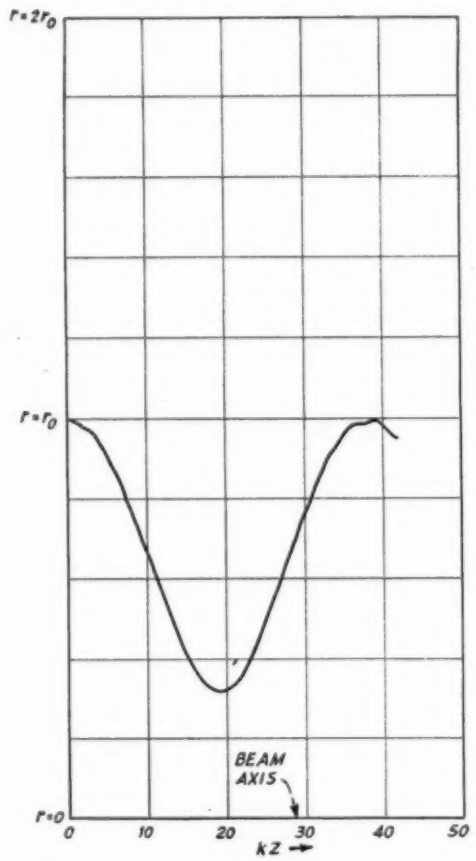


FIG. 7(g). Periodic axially symmetric magnetic focusing. Plasma frequency too low. $A=0.0056$ $B=0.0159$.

The next stable zone appears at

$$\omega_c/a=3.76\omega^2/v. \tag{73}$$

In this and in the higher order stable regions, the electrons are alternately focused and defocused to such a great extent that they travel from one edge of the beam across the axis to the other edge one or more times while traversing a magnet period. None of these higher stable regions represent near-parallel flow.

One of the two basic equations of motion, Eq. (52), has been set up on the laboratories' analog computer and the trajectories of the outer electrons plotted for the case of zero initial transverse velocity. These plots show that the balance conditions as derived in Eqs. (66) and (67) are correct to within a few percent.

Computer solutions corresponding to essentially parallel flow have not been attempted for entrance conditions other than zero transverse velocity and entrance at a maximum of magnetic field. If either of these restrictions is relaxed, there is no simple approximate solution. In this case, when the form of the beam boundary variations is unknown, the assumption of constant space charge force (Eq. (48)) becomes less tenable. For this reason and because it was felt that magnetic quadrupole focusing was of lesser immediate practical importance, the determination of optimum entrance conditions and the description of the beam boundary under these conditions was not pursued.

C. AXIALLY SYMMETRIC PERIODIC ELECTRIC FIELD

Let us now consider the motion of a beam of electrons along the axis of an axially symmetric periodic electric field. This field might be produced by a series of annular disks held at alternately higher and lower potentials as shown in Fig. 14.

Let us call the potential on the axis produced by the electrode system in Fig. 14, $V(z)$ which will be given effectively by

$$V(z)=V_0+V_1(1-\cos kz), \tag{74}$$

provided the origin is at a position of maximum or minimum potential. We have as before, $k=2\pi/p$. With a potential that varies along the axis, we will not have a nearly constant axial velocity for the electrons as before, We will however assume that the electrons have a constant velocity in any given cross section. We will also make assumptions (1) and (2) of Section A. Suppose that r is the radius of an outermost electron, and let

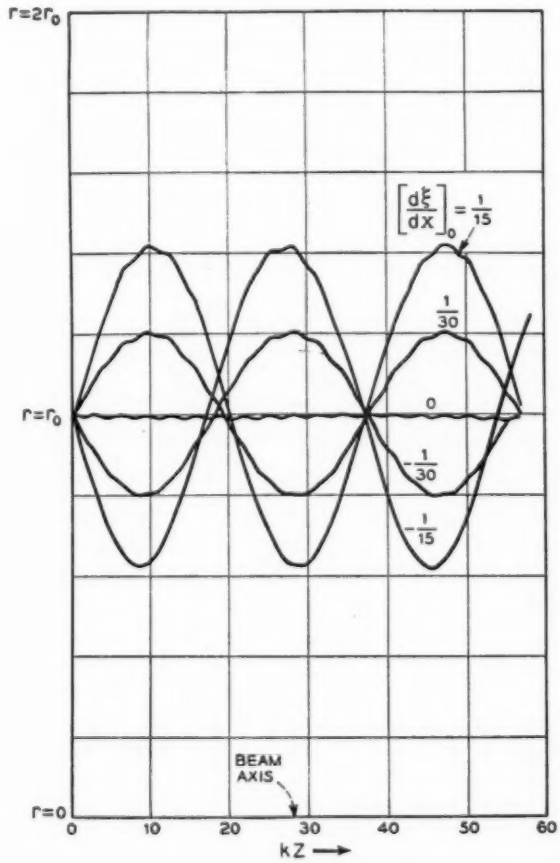


FIG. 8. Periodic axially symmetric magnetic focusing. Effect of radial velocity at beam entrance. $A=0.0139$ $B=0.0139$ $[d\xi/dx]_0=\pm 1/30;\ \pm 1/15$.

$kr=y$. At the origin we have $kr_0=y_0$. Furthermore, let $kz=x$. We can then describe the orbits of the outer electrons by the paraxial ray equation in the form

$$\frac{1}{y_0}\frac{d^2y}{dx^2}+\left(\frac{1}{2V}\right)\frac{1}{y_0}\frac{dy}{dx}+\left(\frac{1}{4V}\frac{d^2V}{dx^2}\right)\frac{y}{y_0}$$
$$-\frac{1}{2}\left(\frac{\omega_p}{\omega}\right)^2\left(\frac{V_0}{V}\right)^{\frac{1}{2}}\frac{y_0}{y}=0. \quad (75)$$

In accordance with assumption (2), we suppose that $(1/y_0)(dy/dx)$ is of first order of smallness. We will also assume that V_1 is small compared to V_0 so that $(1/V)\times(dV/dx)$ is also small in first order. In addition, let (ω_p/ω), y, and y_0 be of first order smallness. Then Eq. (75) is correct to terms in second order provided that y/y_0 is not larger than 10 and not smaller than 10^{-3}. As before, ω_p is the plasma frequency defined by the average charge density at $z=0$, and ω is kv_0 where v_0 is the velocity at $z=0$.

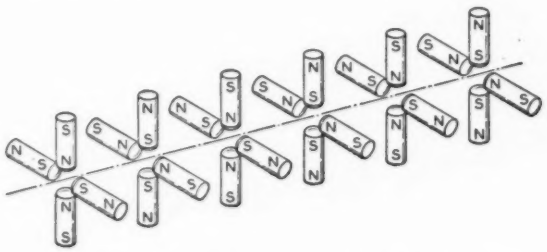


FIG. 9. A magnetic arrangement to give periodic quadrupole fields.

Let us now enter Eq. (75) with the substitutions

$$\xi=y/y_0, \quad (76)$$
$$\Phi=V/V_0, \quad (77)$$
$$A=\tfrac{1}{2}(\omega_p/\omega)^2. \quad (78)$$

We then obtain

$$\frac{d^2\xi}{dx^2}+\left(\frac{1}{2\Phi}\frac{d\Phi}{dx}\right)\frac{d\xi}{dx}+\left(\frac{1}{4\Phi}\frac{d^2\Phi}{dx^2}\right)\xi-\Phi^{-\frac{1}{2}}\frac{A}{\xi}=0, \quad (79)$$

where

$$\Phi=1+\epsilon(1-\cos x) \quad (80)$$

if

$$\epsilon=V_1/V_0.$$

Equation (79) is much more involved than the corresponding Eqs. (10) and (52) with which we discussed the two cases of magnetic focusing. Very fortunately there exists a change of dependent variable that very considerably simplifies affairs.[7] Let us place

$$\sigma=\xi\Phi^{\frac{1}{2}}. \quad (81)$$

[7] J. R. Pierce, *Theory and Design of Electron Beams* (D. Van Nostrand Company, Inc., New York, 1949), p. 85.

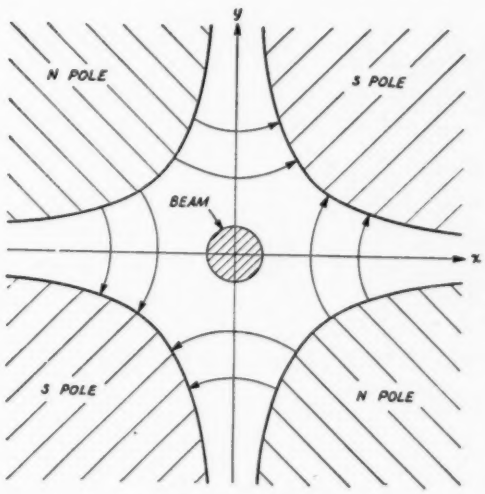


FIG. 10. Quadrupole configuration.

Then we obtain

$$\frac{d\sigma}{dx}=\frac{d\xi}{dx}\Phi^{\frac{1}{2}}+\tfrac{1}{4}\xi\Phi^{-\frac{1}{2}}\frac{d\Phi}{dx} \quad (82)$$

and

$$\frac{d^2\sigma}{dx^2}=\Phi^{\frac{1}{2}}\left[\frac{d^2\xi}{dx^2}+\frac{1}{2\Phi}\frac{d\xi}{dx}\frac{d\Phi}{dx}+\frac{1}{4\Phi}\frac{d^2\Phi}{dx^2}\xi\right]$$
$$-\tfrac{3}{16}\xi\Phi^{-7/4}\left(\frac{d\Phi}{dx}\right)^2. \quad (83)$$

Consequently, we find

$$\frac{d^2\sigma}{dx^2}+\frac{3}{16}\left(\frac{1}{\Phi}\frac{d\Phi}{dx}\right)^2-\frac{A}{\Phi\sigma}=0. \quad (84)$$

Suppose that the electrons start off at $z=0$ in parallel flow so that $(d\xi/dx)_0=0$. We find then from Eqs. (81) and (82) that $\sigma_0=1$ and $(d\sigma/dx)_0=0$.

Let us substitute Eq. (80) into Eq. (84). We obtain to second order

$$\frac{d^2\sigma}{dx^2}+\tfrac{3}{32}\epsilon^2(1-\cos 2x)\sigma-\frac{A}{\sigma}=0. \quad (85)$$

An approximate solution for σ can now be found by placing

$$\sigma=1+\delta(1-\cos 2x). \quad (86)$$

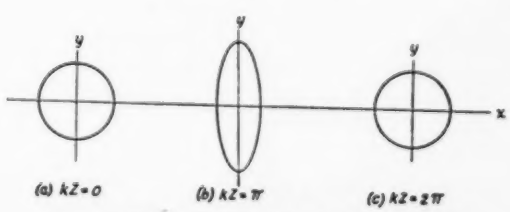


FIG. 11. Cross section of the beam at various distances along axis.

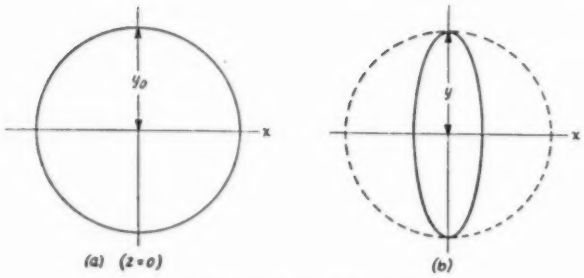


FIG. 12. Comparison of beam shapes.

We find that such an approximation exists with $\delta=A/2$ if $\epsilon^2=(32/3)A$. In other words,

$$\sigma=1+\frac{1}{4}\left(\frac{\omega_p}{\omega}\right)^2(1-\cos 2x) \tag{87}$$

provided

$$\frac{V_1}{V_0}=\frac{4}{\sqrt{3}}\left(\frac{\omega_p}{\omega}\right). \tag{88}$$

Returning to Eq. (81) we find for ξ, to first order,

$$\xi\equiv\frac{r}{r_0}=1-\frac{V_1}{4V_0}(1-\cos x). \tag{89}$$

The essence of the above discussion is that an electron beam satisfying the various limiting conditions that have been imposed can be focused into a beam of nearly parallel flow by a periodic, axially symmetric electric field provided that Eq. (88) is satisfied, and provided the electrons enter the focusing field at a point of zero longitudinal gradient with no transverse motion.

In order to gain some idea of the magnitudes involved let us take a specific case and determine the balance conditions for near-parallel focusing. Assume we have a 1000-volt beam having a diameter of 0.080 in. and carrying a current of 10 ma. Further, let us assume that the periodicity of the focusing apertures is $\frac{1}{4}$ inch. Under these conditions we find

$$\omega_p=1.31\times 10^9,$$
$$\omega=1.14\times 10^{11};$$

then from (88)

$$\frac{V_1}{V_0}=2.65\times 10^{-2}$$

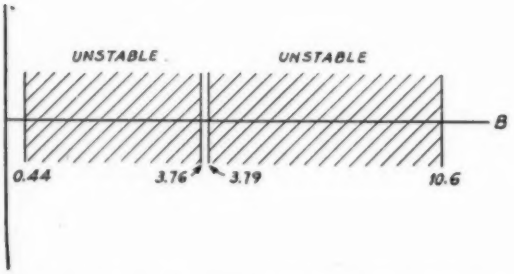


FIG. 13. Regions of stability for quadrupole focusing.

or

$$V_1=26.5 \text{ volts},$$

and from (89) we see that the beam ripples peak-to-peak are about $2\frac{1}{2}$ percent.

D. QUADRUPOLAR PERIODIC ELECTRIC FIELD[8]

As a final case, let us consider the motion of a beam of electrons along the axis of a quadrupolar periodic electric field. This field might be produced by a series of hyperbolic electrodes arranged as shown in Fig. 15.

In a plane perpendicular to the axis the fields will appear as shown in Fig. 16.

It will be noticed that it has been convenient to rotate the coordinate system 45° from the orientation used in the magnetic quadrupole case.

Just as in Sec. C, we can write down an expression for the electric potential and then obtain the electric fields after neglect of terms of order $(kr)^2$. One obtains

$$E_x=\frac{2\Delta V}{a^2}x\cos kz, \tag{90}$$

$$E_y=-\frac{2\Delta V}{a^2}y\cos kz, \tag{91}$$

$$E_z=-k\frac{\Delta V}{a^2}(x^2-y^2)\sin kz, \tag{92}$$

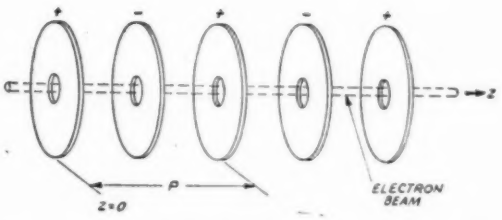


FIG. 14. An electrode configuration to give a periodic axially symmetric electric field.

where ΔV is the magnitude of potential difference between the axis and the electrodes at $z=0$.

It will be immediately noted that the z-directed electric field is zero on the axis. This is quite different from the case of the axially symmetric electric field, and means that we may assume the z-directed velocity v to be constant along the beam.

Let us now make the same assumptions about the space charge forces as was done in the magnetic quadripole case. The motion of an outer electron at $y=0$ is then given by

$$\frac{d^2x}{dt^2}+\left(\frac{2\eta\Delta V}{a^2}\cos kz\right)x-\frac{\omega_p^2}{2}x_0=0, \tag{93}$$

[8] This form of focusing has been described by J. P. Blewett, Phys. Rev. **88**, 1197–1199 (1952) for the case where space-charge defocusing is negligible.

while the motion of an outer electron at $x=0$ is described by

$$\frac{d^2y}{dt^2}-\left(\frac{2\eta\Delta V}{a^2}\cos kz\right)y-\frac{\omega_p^2}{2}y_0=0. \tag{94}$$

Let us now, as before, place $t=z/v$, $kz=q$, $\omega=kv$, $\xi=x/x_0$, and $\zeta=y/y_0$. Equations (93) and (94) then become

$$\frac{d^2\xi}{dq^2}+\left(\frac{2\eta\Delta V}{\omega^2a^2}\cos q\right)\xi-\frac{1}{2}\left(\frac{\omega_p}{\omega}\right)^2=0, \tag{95}$$

$$\frac{d^2\zeta}{dq^2}-\left(\frac{2\eta\Delta V}{\omega^2a^2}\cos q\right)\zeta-\frac{1}{2}\left(\frac{\omega_p}{\omega}\right)^2=0. \tag{96}$$

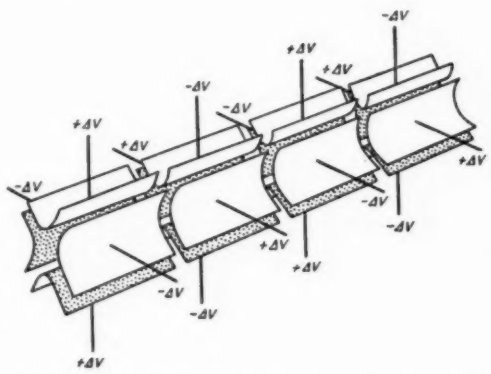


FIG. 15. Hyperbolic electrodes to produce periodic quadrupolar electric field.

Let us as usual place $A=\frac{1}{2}(\omega_p/\omega)^2$. Let us also define

$$B=\frac{2\eta\Delta V}{\omega^2a^2}=\frac{1}{(ka)^2}\frac{\Delta V}{V}, \tag{97}$$

where V is the constant axial potential. The equations of motion now become

$$\frac{d^2\xi}{dq^2}+(B\cos q)\xi-A=0, \tag{98}$$

$$\frac{d^2\zeta}{dq^2}-(B\cos q)\zeta-A=0. \tag{99}$$

We have seen in Sec. B that these equations can simultaneously describe nearly parallel flow only if

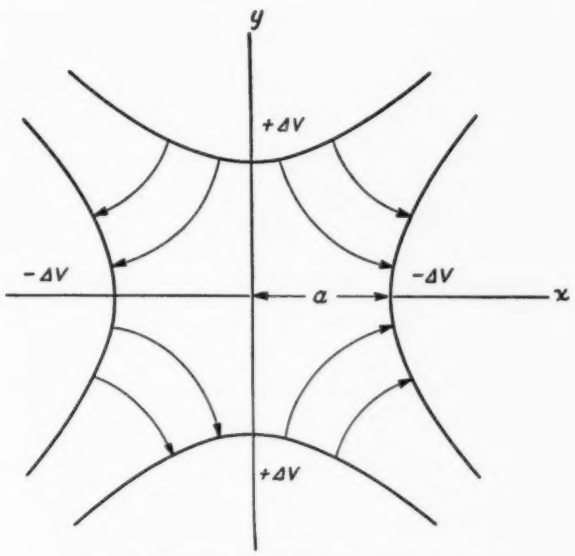


FIG. 16. Quadrupole field configuration in the transverse plane.

$B\ll1$. If we assume that this is the case, the condition for such a motion is

$$B=(2A)^{\frac{1}{2}} \tag{100}$$

and the outer orbits are given by

$$x=x_0[1-\epsilon(1-\cos kz)], \tag{101}$$

$$y=y_0[1+\epsilon(1-\cos kz)], \tag{102}$$

where

$$\epsilon=(2A)^{\frac{1}{2}}. \tag{103}$$

Using Eq. (97), Eqs. (100) and (103) can be written

$$\Delta V/V=(ka)^2(\omega_p/\omega), \tag{104}$$

$$\epsilon=(\omega_p/\omega). \tag{105}$$

As before, the first unstable zone appears at $B=0.44$, which corresponds to a value of $\Delta V/V$ given by

$$\Delta V/V=0.44(ka)^2. \tag{106}$$

The authors wish to acknowledge the many contributions of their colleagues to this paper and to mention in particular J. T. Mendel, W. H. Yocom, J. C. Ward, C. F. Quate, and R. Kompfner who have been concurrently pursuing experimental and theoretical work on this subject. R. W. Hamming has been most helpful in directing the work done on the analog computer.

Germanium Diodes from Spherical Pellets*

W. C. DUNLAP, JR.
General Electric Research Laboratory, Schenectady, New York
(Received August 17, 1953)

A method is described for constructing germanium high-voltage diodes, utilizing 15-mil germanium spheres instead of the usual flat wafer. The spheres are produced within a few minutes in quantities of several hundred thousand or more by blowing molten high-purity germanium from a graphite crucible. The spherical pellets can be annealed, ground, etched, and assembled into diodes by techniques easily adaptable to automatic mass production. The assembled diode uses a sphere-plane contact rather than the conventional whisker contact. Peak back-voltages in the range 50–100 volts are easily obtainable. Series-assembled units for higher voltages are easily made.

INTRODUCTION

COMMERCIAL germanium diodes are made, at the present time, much as shown in the diagram of Fig. 1. A small wafer of germanium, sawed out of a larger ingot, is soldered to a pin, and a fine whisker is brought into contact with the etched surface of this wafer to make the point contact. The present paper deals with attempts to simplify design and production of these diodes by replacing the wafer by a spherical pellet, and by replacing the whisker by a flat plate. These changes may make possible great simplification of the problem of automatic mass production of germanium diodes.

Prior to the present work, spherical pellets of germanium were made by H. Q. North.† The pellets he produced were made by forcing molten germanium out of a graphite crucible by means of a graphite piston. These pellets were about 60 mils in diameter and were of low quality for diode work.

The present studies were based upon the idea that ejection of the pellets by a pure gas would lead to superior characteristics because of the reduced contamination. This hope was actually borne out by the results, and it has been found possible to make diodes, approximating those of commercial production, with the spherical pellet in contact with a plane, rather than with a pointed whisker.

Following a discussion of methods for production, handling, and application of the spherical pellets, an evaluation is given of the possibilities of the sphere-plane contact for single diodes. Some discussion is also given of the use of stacked assemblies of sphere-plane diodes for high-voltage applications (kilovolt range).

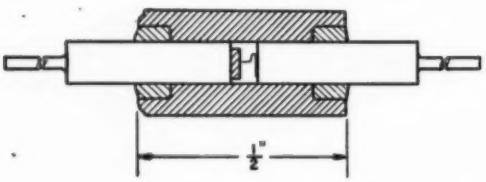


FIG. 1. Conventional germanium diode longitudinal section. The germanium pellet is a plate sawed out of a germanium ingot, ground, polished, and etched. The whisker is a very fine, pointed wire, usually platinum or tungsten, sharpened, bent, and welded to the pin.

PRODUCTION AND PROPERTIES OF 15-MIL GERMANIUM SPHERES

Figure 2 shows a diagram of the pellet-blowing apparatus, and Fig. 3 a picture of the crucible and attachments. The process of making the pellets is quite simple: highly purified germanium of *n* type (30–40 ohm cm) is put into the crucible, attachments are made, water, gas, and electrical power are turned on,

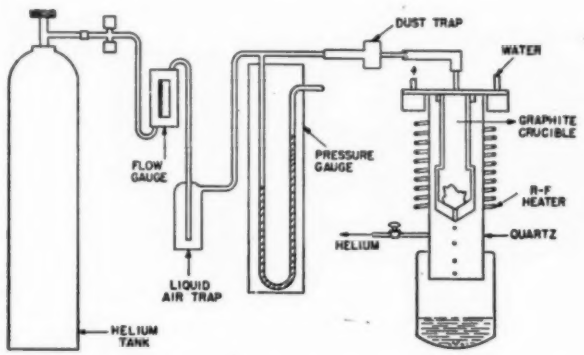


FIG. 2. Schematic diagram of apparatus for blowing germanium spheres. The gas ejects molten germanium from the crucible almost as fast as it melts off the solid ingot. There is relatively little contamination by the crucible, which is also highly purified.

and within a few minutes one can produce automatically several hundred thousand or more pellets.

It is seen from Fig. 2 that the production of the pellets is accomplished inside a quartz jar. The graphite parts screwed into a molybdenum holder, which was water-cooled to reduce contamination of the graphite and the germanium by the metal. Use of the radio-frequency heating method made possible rapid heating of the graphite and reduction of gas coming from the quartz walls as a source of contamination.

The gas used in the present work was helium, although other inert gases can be used. The gas was dried over a liquid nitrogen trap before passing it through

* A preliminary discussion of these studies was given at the Institute of Radio Engineers Electron Devices Conference at the University of Michigan, Ann Arbor, June, 1950.
† Unpublished work.

the system. The connections were made tight so that the gas must pass through the 15-mil orifice at the tip of the graphite crucible.

When the rf heater was turned on, the graphite heated up rapidly. As the germanium became hot enough to melt, liquid germanium ran off the ingot and filled up the crucible. If the gas pressure is great enough to overcome the surface tension at the 15-mil orifice, the germanium leaves the crucible in a continuous stream of droplets, which fall into the bottom of the jar, usually into oil or water. Rapid quenching takes place there, and usually *p*-type pellets are produced. Annealing at 500°C for one day leads to conversion to *n* type of practically all the pellets.

It was found that if the gas pressure were only slightly above the minimum required to force the geramnium out, droplets tended to merge with each other on leaving. This led to a distribution of pellet sizes. With excess pressure, the pellets were nearly uniform in size and shape. For a 15-mil orifice, a pressure of 9

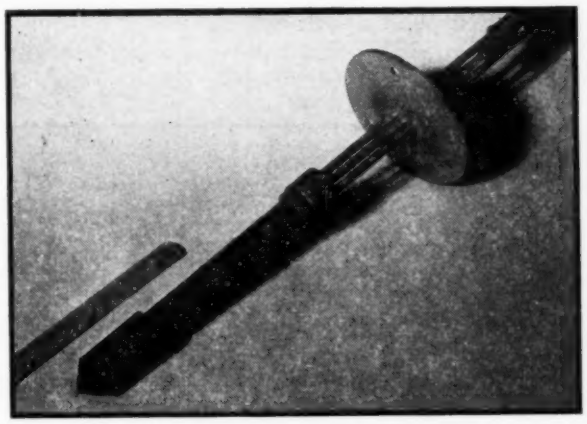


FIG. 3. Photograph of the crucible assembly used in the present work, including the water-cooled molybdenum attachment.

cm Hg was needed to eject the droplets. A pressure of 20 cm Hg was found adequate for uniform particle size.

The graphite used was high-purity Grade D material especially purified for this work by firing at a temperature of about 1800°C in vacuum for several hours.

Figure 4 shows a group of 15-mil pellets produced by the above methods. They are seen to be not completely spherical. Corrugations, points, and other irregularities can be seen. These result from the expansion of the germanium on freezing, and from various crystal growth factors such as grain boundaries, lineage, etc.

X-ray studies made on some of the pellets showed that the majority contained several crystals. This is a drawback for diode work. With more attention to slow cooling of the droplets and modification of the quench step, better crystal quality might be obtained.

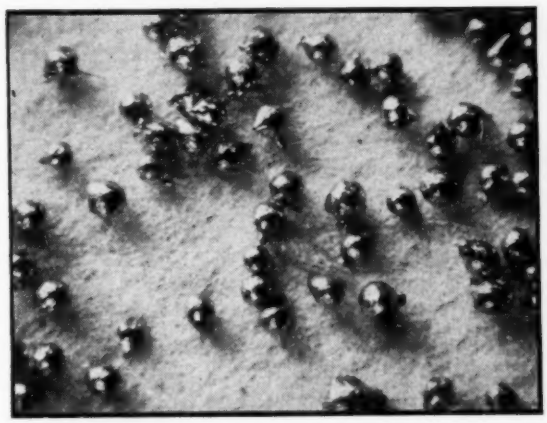


FIG. 4. Photograph at 50× amplification of 15-mil germanium pellets.

HANDLING AND APPLICATION OF THE SPHERICAL PELLETS

As seen from the foregoing, large numbers of germanium pellets of convenient size and shape can be produced by the present method. It is found that every 100 g of germanium yields about 300 000 pellets, produced automatically in a few minutes. These pellets are usable with little further handling, and there is no loss of material as a result of sawing.

In general, there are other operations that one wishes to make on a germanium pellet before assembling the device itself. Thus, we may wish to grind, polish, or etch the pellets to make them more uniform, remove surface irregularities, remove surface contamination, etc. The spherical pellet process is distinguished by the great ease with which mass operations on the pellets can be performed. Thus, polishing and grinding have been done on large numbers at one time by tumbling the pellets with the proper abrasives in a glass jar. Etching of the pellets can also be carried out *en masse* in a similar fashion.

For making a germanium diode, it is necessary usually to solder the pellet to a pin. Extremely small diodes, smaller than the head of a pin, have been made with the 15-mil pellets, by soldering the sphere to a 15-mil wire, contacting with another piece of wire, and sealing in a droplet of resin. However, for reasonable current capacity, it is advisable to use larger pins, and we have soldered generally to pins in the range 40–80 mils in diameter.

For handling operations, either automatic or by hand, vacuum pipettes are recommended. It is desirable to etch lightly and rinse with distilled water following the soldering. In general, the same precautions must be observed for handling spherical-pellet units as are used for conventional diode construction.

SPHERICAL-PELLET DIODES

Because of the large curvature of the surface of the 15-mil pellet, and of the existence of slight surface irregularities and points, it seemed logical to use the

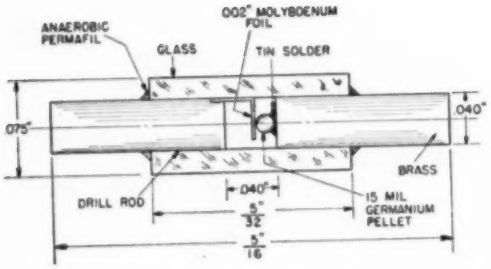


FIG. 5. Scale drawing of spherical pellet diode embodying the sphere-plane contact. A ribbon of molybdenum or nickel has been used for the planar contact.

germanium pellet as the source of the point contact, and to replace the usual pointed whisker with a flat surface. One design for accomplishing this is shown in Fig. 5. All the diodes discussed here were made with the sphere-plane contact. The replacement of the pointed whisker offers a new means of simplification of the diode construction problem.

Figure 6 shows several completed germanium diodes made from spherical pellets, with a conventional diode included at the top, for comparison. The power-handling capacities of all the diodes shown were about the same, in spite of the different sizes.

Some of the diodes made during this study were made as shown in Fig. 5. Others simply involved use of a solid cylindrical pin for the contact. The ribbon construction is best suited for single diode construction, since the solid-pin contact is highly rigid and contact is often lost due to expansion of the case. For the series-connected assemblies, separate springs were used, and the solid-pin construction was found satisfactory.

Figure 7 shows the voltage-current characteristic of a typical unpulsed spherical-pellet diode. The reverse current is seen to be quite large. The forward characteristic is acceptable. It appears that, in the present preliminary stage of development, the spherical-pellet diodes are suitable for relatively low-voltage applications where fairly large reverse currents can be tolerated.

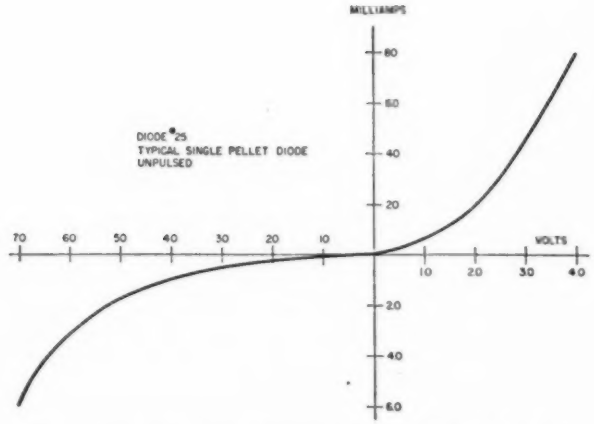


FIG. 7. Current-voltage characteristic of a spherical-pellet diode with planar contact.

The reversed characteristics were considerably improved by ac pulsing, although the above statements still remain true in general. The best pulsing treatment involved passing a 30-volt, 60-cycle pulse through the

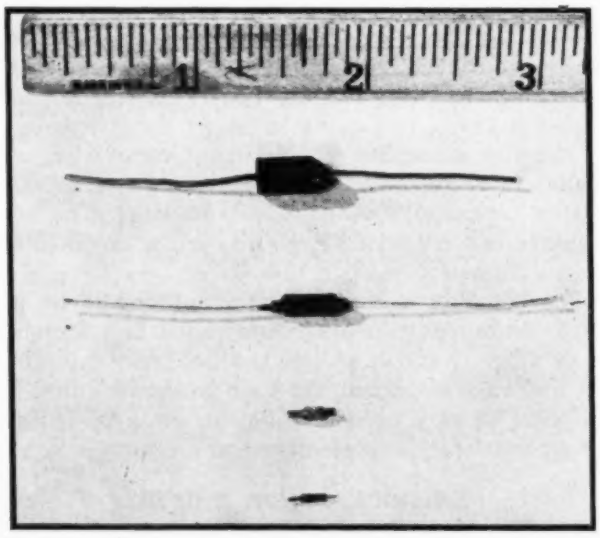


FIG. 6. Spherical-pellet diodes compared to a conventional commercial diode (at top). All these diodes had about the same current capacity and peak-back voltage.

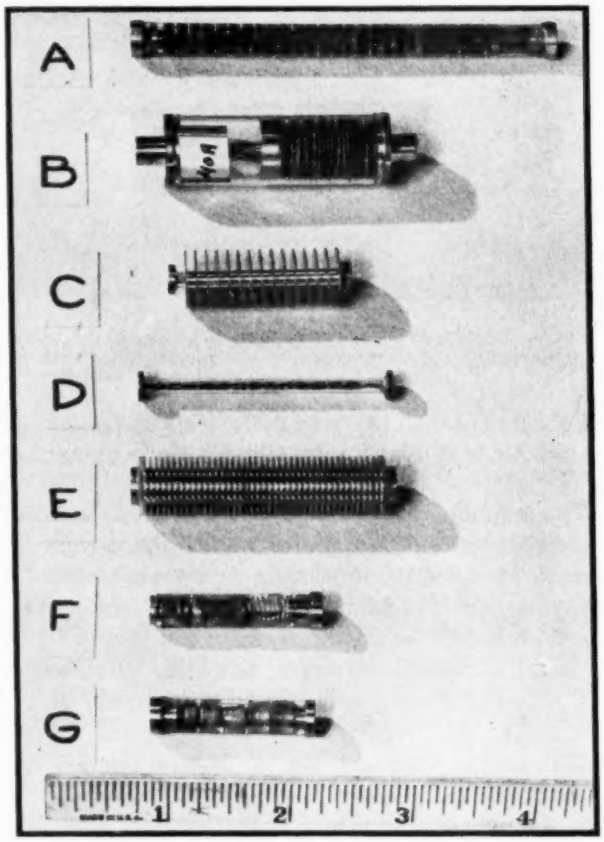


FIG. 8. Photograph of a series assemblies of stacked pellet-diode units. Here the base of the pin, to which each sphere is soldered, acts as the contactor for the next sphere. Back-voltage rating ranged up to 1000 volts.

diode, through a 50-ohm resistor, for a time of 0.8 second. For series-connected units, 30 volts was added for each added unit.

Several hundred diodes have been made and tested from various pellet runs. There was considerable spread in characteristics, and peak voltage ratings ranged from 20 to over 120 volts. For a typical group of 25 completed diodes, the following average characteristics were obtained: (a) peak-back voltage, 71 volts; (b) reverse current at 40 volts, 2.7 ma; (c) forward current at 1 volt, 7.1 ma.

ASSEMBLIES OF SPHERICAL PELLET DIODES

A number of rectifiers were made containing spherical-pellet diodes stacked in series. An example is shown in Fig. 8. Here the pins were made broad, sometimes very short, so that as many as 25 units could be stacked per inch of length. The units were usually assembled inside a glass tube, with a spring at the end to keep the contact. Several 1000-volt assemblies were made using 25 units. The characteristics are quite similar to those of the single units, with the following qualifications: (a) Because of the effect of each diode on others close by, the reverse characteristics are poorer than for the single units. A group of diodes testing 60–70 volts peak inverse when tested singly, average usually only about 40 volts per diode when made into an assembly before testing. (b) The forward characteristic deteriorates in that it takes about 1 volt per diode unit to produce a forward current of a few ma. (c) Heat dissipating fins have been incorporated into several of the assemblies, and these do lead to a partial restoration of the characteristics deteriorated by thermal effects.

SUMMARY AND CONCLUSIONS

In the above, we have discussed the preparation and properties of germanium spheres prepared by ejection of molten germanium droplets from a graphite crucible under pressure of a gas. We have shown that rectifiers acceptable for certain applications can be made from these pellets. We have pointed out certain advantages accruing from the use of this method: great ease in the production and handling of the pellets to be used for diodes and other semiconductor devices. We have also discussed replacement of the usual whisker contact of diodes by a sphere-plane contact, which is a logical application of the use of very small, 15-mil spherical pellets. It has been pointed out, and should be emphasized here, that, compared to the best rectifiers and diodes now being made of germanium, the present devices are relatively unspectacular and that the question becomes one of resolving the quality needed relative to the costs of production. For many applications, it is believed that the probable low cost of the spherical-pellet units, and their adaptability to mass production, may give them a certain interest to engineers and production designers in this field.

ACKNOWLEDGMENT

It is desired here to acknowledge the substantial contributions to this work made by Mr. G. T. Jacobi, Mr. John Marinace, and Mr. Harold Rochford.

JOURNAL OF APPLIED PHYSICS VOLUME 25, NUMBER 4 APRIL, 1954

Studies on the Creep Recovery and Annealing of Zinc Single Crystals*

M. TANENBAUM † AND W. KAUZMANN
Frick Chemical Laboratory, Princeton University, Princeton, New Jersey

(Received September 8, 1953)

Part I is a description of some of the factors involved in the recovery by pure zinc single crystals at 35°C of the ability to creep following plastic deformation. The following observations were made: (1) Complete recovery of the ability to show an initial instantaneous plastic extension occurs much less readily than recovery of the ability to show transient creep. (2) Once the quasi-viscous creep stage is reached no recovery of transient creep occurs when the load is partially or completely removed for two minutes, but removal of only 6 percent of the load for 100 minutes permits partial recovery of transient creep at the full load. A 300 minute "rest" at 91.5 percent of the full load resulted in complete recovery of the transient creep at full load, even though creep occurred at the reduced stress. (3) The recovery of quasi-viscous creep at a reduced stress after exposure to a higher stress was also investigated. The recovery was slower, the larger the initial load. (4) The rate of the time-dependent contraction of zinc crystals on removal of the load probably follows a power law in the time. There is also an instantaneous nonelastic contraction which is about ten times larger than the elastic contraction.

These results show that the role of thermal fluctuations in activating the units of flow is very different from the role of the stress. The applications of the Eyring rate equation to quasi-viscous creep which have been made in the past are, therefore, not valid. A variable activation energy must be postulated in order to account for the kinetics of creep and recovery. These results are interpreted in terms of the dislocation theory of plastic flow.

In Part II a phenomenon is described in which pure zinc single crystals are consistently found to become temporarily hardened by annealing in vacuum above 200–260°C. The hardness thus introduced can be removed by straining the crystal and then allowing it to stand at temperatures below 200°C. The crystal is in this way returned to its normal plastic state. The hardening is not removed by electropolishing the crystal, so is not a surface phenomenon. Though similar to Orowan's thermal hardening effect, it differs by requiring a much higher temperature and by not requiring the addition of impurities.

* Taken from a thesis submitted by M. Tanenbaum in partial fulfillment of the requirements of the degree of Doctor of Philosophy at Princeton University.

† Present address: Bell Telephone Laboratories, Murray Hill, New Jersey.

IF a single crystal of zinc is stressed beyond the elastic limit, the resulting strain can be resolved into the following elastic and plastic components: (1) An instantaneous elastic strain which is proportional to the applied stress. On removing the stress, the elastic strain is instantaneously and completely recovered. (2) A very rapid, if not instantaneous, plastic strain which is not proportional to the applied stress ("initial plastic strain"). Most of the initial plastic strain is not recovered when the stress is relaxed. (3) A time-dependent plastic deformation whose rate decreases with time ("transient creep"). (4) Following transient creep a stage of deformation may be observed in which the rate of strain is constant with time ("quasi-viscous" or "steady-state" creep). (5) A stage of creep may also be observed in which the strain rate begins to increase with time ("tertiary creep").

If the crystal is plastically strained, unloaded, and then immediately reloaded, it is generally found that the elastic limit has been increased, and the transient creep rate is less than that under the same stress in the previous test. This phenomenon is known as work-hardening. If, however, the crystal is annealed at a suitable temperature for a sufficient time, the work hardening disappears, the elastic limit drops to its former value, and the creep behavior observed in the first test can usually be reproduced. The removal of work-hardening is known as "recovery." Crystals of zinc can recover at room temperature if given sufficient time, but the recovery is greatly accelerated by raising the temperature.

The purpose of the work reported in the first part of this paper was to study some of the variables which determine the recovery of zinc single crystals at 35°C. in order to obtain a better understanding of the fundamental processes which are involved. In the second part we shall describe a rather unexpected phenomenon which takes place on annealing pure zinc crystals at elevated temperatures.

EXPERIMENTAL

Preparation of the Crystals

The crystals were grown from 99.999+ percent zinc supplied by the New Jersey Zinc Company. Crucibles were prepared by drilling a $\frac{1}{4}$ in. hole axially in a $\frac{1}{2}$ in. rod of spectroscopically pure graphite. The zinc used in the first few crystals (I–VI) was degassed and cast into rods by the method of Slifkin and Kauzmann.[1] The crucible containing its charge of zinc was then placed in a glass tube which was evacuated to 0.01 micron and sealed off. The crystal was grown by heating the sealed tube above the melting point of the zinc and then slowly lowering it through a sharp temperature gradient at a rate of 3.3 cm/hour. A much more efficient degassing technique was later developed in which the zinc was first sublimed into a Pyrex vessel and then melted into the crucible, the entire process being carried out under a vacuum of better than 0.1 micron. The zinc used in making crystals VII–XXII was degassed in this way.

The orientations of the crystals were determined optically either by the cork-ball method of Bridgman[2] or by a two-circle goniometer. Etch pits were developed by an acidic $CuCl_2$ solution described by Barrett.[3] The cork-ball method permits an error of about 2 degrees. The goniometer reduced this error to less than 1 degree.

For purposes of loading, grips were fastened to the ends of the crystals. These grips were designed to accommodate the crystal as grown, without necessitating any trimming of the ends. The grips were cemented to the crystal with a rubber-base cement (Core Cement No. 121-Xylos Division, Firestone Tire and Rubber Company, Akron, Ohio) which resisted temperatures as high as 400°C with a minimum of cracking.

A pair of sleeves were attached to the crystal at least $\frac{1}{4}$ in. from the grips. In most of the experiments they consisted of Vycor glass tubing whose surfaces were roughed on a carborundum wheel. The sleeves were cemented to the crystal with Firestone Core Cement No. 121 and served both to determine the gauge length of the sample (usually 8 to 10 cm) and to provide surfaces on which the extensometers could be mounted. It was later found expedient to employ a pair of split sleeves machined from aluminum which screwed together tightly around the crystal $\frac{1}{4}$ in. or more from the grips and which could be readily removed without damage to the specimen. These sleeves were used on crystals XVI–XXII.

The extensometers were of the optical lever type, consisting of mirrors attached to rollers which moved between two parallel, optically flat agate surfaces connected to the two sleeves. The rollers were pieces of $\frac{1}{16}$ in. drill rod carefully polished and circular in cross section to within 0.0002 in. Reflections of scales 5.6 meters away were observed in the mirrors through 20 power telescopes. With a gauge length of 8 cm, strains of 40A/cm could be measured.

Two extensometers were mounted on opposite sides of each specimen and their readings were averaged to cancel out any errors caused by bending of the crystal. They were clamped to the specimen by springs stretched between them and their weight was counterbalanced by springs attached to the upper grip of the specimen. The extensometers were tested by determining the elastic modulus of mild steel, values in excellent agreement with accepted ones being obtained.

The loading frame by which the cyrstals were stressed and the mechanisms for temperature control are described by Slifkin and Kauzmann.[1] All experiments were performed at 34.25°±0.02°C.

[1] L. Slifkin and W. J. Kauzmann, J. Appl. Phys. **23**, 746 (1952).

[2] P. W. Bridgman, Proc. Am. Acad. Arts Sci. **58**, 165 (1923); **60**, 305 (1925).

[3] C. S. Barrett, Trans. Am. Inst. Mining Met. Engrs. **124**, 29 (1937).

Annealing was performed at a pressure of less than 1 mm of Hg. Below 260°C the annealing was carried out in a Pyrex vessel with a 102/75 ground-glass ball-and-socket joint which accommodated both the extensometers and the specimens. Above 260°C a vessel was employed which necessitated removal of the extensometers from the specimen. Sleeve clamps were developed which held the specimen immobile and allowed removal and remounting of the extensometers without bending the specimen.

With the techniques employed, creep rates could be measured with an error of about 1×10^{-6} minute^{-1} when readings were taken every minute and with correspondingly greater precision over longer time intervals. The error in calculating the resolved shear stress in the slip direction was about 2 percent.

Throughout this paper creep rates will be expressed in terms of resolved shear strain rates unless otherwise stated. In the hexagonal crystal lattice of zinc, the elongation ϵ is related to the shear strain in the slip direction, s, by the equation

$$\epsilon = s \sin\alpha \cos\beta,$$

where α is the angle between the C axis and the specimen axis and β is the angle between the projection of the specimen axis on the basal plane and the nearest slip direction. Since α was usually 40 to 85 degrees and β cannot be greater than 30 degrees, ϵ and s seldom differed by more than a factor of three. In comparing runs on a given crystal there is no error, since the change in α and β due to lattice rotation during the course of the tests was always too small to affect the creep rate within the limits of experimental error.

RESULTS

Part I. Creep Recovery at 35°C

This work represents the results of about 300 creep tests on eighteen different crystals. The crystals were usually allowed to creep from 60 to 300 minutes at resolved shear stresses of $1\text{–}2\times10^6$ dynes/cm^2. The shear strain produced in a given run was almost never greater than 0.1 percent and the effect of the strain on the creep behavior could invariably be removed by annealing the crystal at 180–200°C for 3–5 hours. This permitted many runs to be made on the same crystal, thus eliminating the uncontrollable variability that is observed between crystals which are grown in an apparently identical manner.

In contrast to the results of Slifkin and Kauzmann,[1] the crystals studied in the present work usually showed appreciable quasi-viscous creep within the first 100 minutes of a creep test. It is, therefore, necessary to describe the plastic behavior at constant stress by the complete Andrade equation[5]

$$\epsilon = \epsilon_0 + bt^m + kt,$$

[5] E. N. da C. Andrade, Proc. Phys. Soc. (London) **84A**, 1 (1910); **90A**, 329 (1914).

TABLE I. The effect of annealing on initial plastic strain.

Crystal	Resolved shear stress dynes/cm^2 $\times10^{-6}$	Annealing condition Temperature °C	Annealing condition Duration hr	$\epsilon_0\times10^4$
V	1.67	25	24	5.2
	1.67	25	190	10.5
	1.67	215	13	15.3
	1.81	25	13	30.0
	1.81	225	22	43.3
XIII	1.960	25	20	14.7
	1.960	180	5	68.8

where ϵ is the total plastic strain at time t, ϵ_0 is the initial plastic strain observed immediately after applying the load, b and m are constants relating to transient creep while k is the quasi-viscous creep rate.

The Initial Plastic Strain

The quantity ϵ_0 has seldom been studied by previous investigators. Slifkin and Kauzmann[1] have reported that ϵ_0 is proportional to b in a given set of runs on a particular crystal. The values of ϵ_0 and b for thirty runs on one of our crystals were compared and the results indicated a similar relationship between the two quantities. The scatter was, however, much greater than that observed by Slifkin and Kauzmann.

It was also found that ϵ_0 is far more sensitive to the past history of the specimen than the transient or quasi-viscous creep rates. Although recovery may have proceeded to the point where b and k indicate the complete removal of work-hardening, the value of ϵ_0 may be far below that for the completely recovered specimen. Table I indicates this sensitivity in the case of three different crystals. For each crystal the transient and quasi-viscous creep rates were the same within the normal range of variations, indicating equal recovery, yet ϵ_0 is much larger after annealing at high temperatures.

Recovery of the Ability to Creep Transiently

Slifkin and Kauzmann[1] found that removal of the load from a crystal for a few minutes during transient creep resulted in recovery of a part of the ability to creep transiently; that is, reapplication of the original load resulted in a higher rate of creep than that observed just before removing the load. In other experiments Slifkin[6] found that no such rapid recovery of transient creep occurred when only part of the load was removed.

Carrying this investigation a step further, a series of experiments was performed to study the recovery of the ability to creep transiently after the crystal had entered the quasi-viscous region. Crystal XIII was stressed at 1.96×10^6 dynes/cm^2 for 100 minutes, at which time various fractions of the load were removed

[6] L. Slifkin, thesis, Princeton University (1950).

for two minutes and then replaced. In this way, the crystal was allowed to rest for two minutes under 90 percent, 80 percent, 70 percent, 40 percent, 20 percent, and 0 percent of the initial load. The strain that occurred in the first twenty minutes after reapplying the load was compared with that which would have been observed had the crystal been allowed to continue to creep quasi-viscously without reducing the load. No significant difference between the two strains was observed. There was also no indication of transient creep upon reapplication of the original load. It can be concluded that once the crystal has reached the quasi-viscous state, no significant recovery of the ability to creep transiently occurs on partial or complete removal of the load for two minutes.

The results of another series of experiments are illustrated in Fig. 1. Crystal XIII was allowed to creep at a resolved shear stress of 1.96×10^6 dynes/cm^2 for 100 minutes. Then the load was reduced by 6 percent by removing a "flyweight" from the loading beam without otherwise disturbing the specimen. The specimen was allowed to "rest" under the lower stress for 100 minutes (during which time slow creep was observed as shown by the lower curve in Fig. 1) and then the stress was returned to its original value and the resulting creep was noted. The crystal was now annealed for 5 hours at 180°C and the experiment was repeated except that the entire load was removed during the 100-minute rest period. Surprisingly enough, it was found that a crystal can recover the ability to creep transiently just as well in 100 minutes at a rest load equal to 94 percent of the initial load as under zero rest load.

Figure 2 shows that a crystal (Crystal V) can fully recover the ability to creep transiently under a rest load equal to 91.5 percent of the test load in a rest period of only 300 minutes, despite the fact that it was creeping at

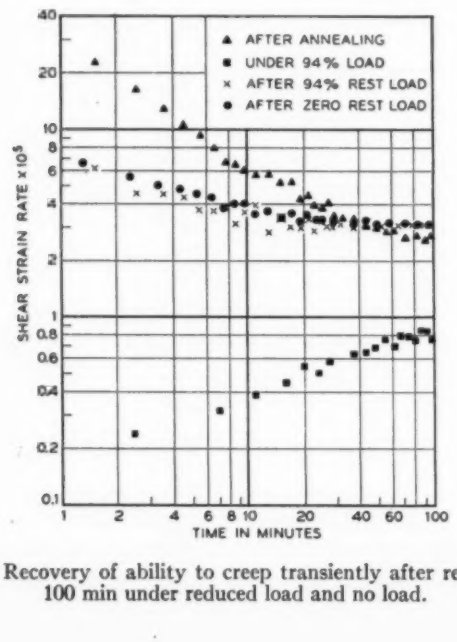


FIG. 1. Recovery of ability to creep transiently after resting for 100 min under reduced load and no load.

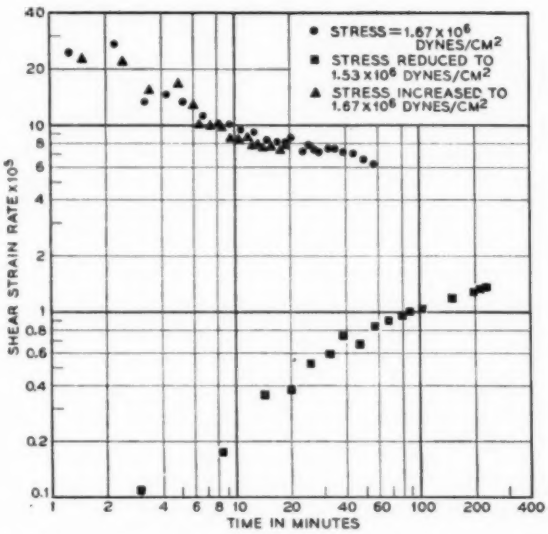


FIG. 2. Recovery of ability to creep transiently after resting for 300 min under reduced load.

an appreciable rate under the rest load. The crystal was stressed at 1.67×10^6 dynes/cm^2 (resulting creep shown by black circles in Fig. 2). After 60 minutes the stress was reduced by 8.5 percent (subsequent creep shown by black squares). The stress was then returned to its original value (resulting creep shown by black triangles). We have not determined how large the rest load may be before it begins to have an appreciable effect on the recovery, but the fact that it may be as large as 91.5 percent of the test load is interesting and significant.

It should also be mentioned that although a large rest load does not seem to interfere with the slow recovery of transient creep, it does have a big effect on ϵ_0. Thus, on returning to the initial load after resting 100 minutes at no load, ϵ_0 was ten times larger than the ϵ_0 obtained after a 100 minute rest at 94 percent of the initial load. This is another illustration of the sensitivity of ϵ_0 to the past history of the specimen.

Recovery of Quasi-Viscous Creep After Reduction of the Load

Consider an annealed crystal which is loaded at stress A and allowed to creep, after which the stress is lowered to a value B. Slifkin[6] found that if this change is made early in the transient creep stage, the rate would fall immediately to the value expected at the same strain had the entire run been made at stress B. In a few exploratory runs on one of his crystals, however, he noted that if he waited until the rate at stress A had leveled off to a constant quasi-viscous value, the reduction of stress to B produced a decrease in rate to almost zero. The rate then slowly increased. It was felt that this interesting result required further study.

An investigation of several crystals has shown that when the load on a crystal is suddenly reduced by 5–10 percent during quasi-viscous creep, the following

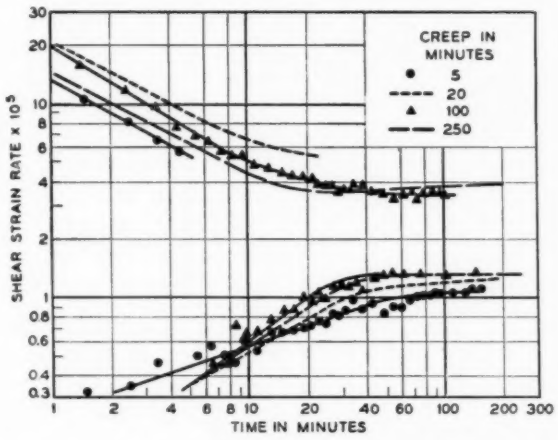


FIG. 3. Creep behavior on reducing load by 6.1 percent after various times at a resolved shear stress of 1.58×10^6 dynes/cm². The upper curves show the creep before reducing the load and the lower curves show the creep after reducing the load.

changes take place: (1) The creep rate immediately falls to a very low value, perhaps to zero. (2) The creep rate then slowly rises, the logarithm of the creep rate being approximately proportional to the logarithm of the time. (3) The rate finally levels off at the quasi-viscous value that would have been obtained had the freshly annealed crystal been originally loaded at the lower stress. (4) If the load is returned to its higher value once the new steady quasi-viscous rate is obtained, the crystal will (with the exception of ϵ_0) appear to be fully recovered and show the same transient creep observed on the freshly annealed crystal at the higher stress.

Figure 3 illustrates the behavior of Crystal V when 6.1 percent of the load was removed after various times of creep at the higher stress. Once the crystal has begun to creep quasi-viscously, the recovery curves under the lower load are the same, no matter how long the quasi-viscous creep had continued before decreasing the load. The only curve showing any appreciable deviation from this behavior is the one obtained after five minutes of creep at the higher stress, where the crystal is still creeping transiently. This fact supports the postulate that quasi-viscous creep represents a true stationary state of the crystal.

Figure 4 illustrates the recovery of Crystal XI at a given load after creeping for 40 minutes at various higher loads. As might have been anticipated, it is found that the higher the initial load, the lower the rate immediately after decreasing the stress.

Contraction After Creep

Slifkin and Kauzmann[1] have reported a contraction of zinc single crystals upon relaxing the stress after plastic deformation that is a factor of 10 greater than the calculated elastic contraction. This contraction consists of an instantaneous component followed by a slow time-dependent contraction. No correlation was obtained between the amount of contraction and the amount of plastic strain that preceded it.

A similar contraction was invariably observed after every creep test in the present work. The instantaneous contraction was usually a factor of 10 smaller than ϵ_0 and about 10 times larger than the calculated elastic contraction. Again no correlation between the creep history of the specimen and the amount of contraction was observed, but in a series of experiments with Crystal XIII in which increments of the load were removed during quasi-viscous creep, the instantaneous contraction was roughly proportional to the fraction of the load removed for fractions between 0.1 and 1.0.

Time-dependent contraction was also observed with every crystal. This phenomenon was not studied in detail since its rate was too close to the limit of sensitivity of the extensometer and long time intervals between readings were necessary to permit any reliable estimation of the rate of contraction. This contraction could, however, be measured during two runs on Crystal XIII, and the results are shown in Fig. 5. It will be noted that we again obtain more or less of a proportionality between the logarithm of the rate and the logarithm of time, at least in one of the runs.

DISCUSSION

We have seen that during quasi-viscous creep, if the stress is reduced by 5–10 percent, the creep rate drops immediately to a very low value and then slowly rises to the quasi-viscous rate expected for the particular crystal at the lower stress. Furthermore, the recovery of the ability to creep transiently at a given stress proceeds just as rapidly at a stress equal to 90–95 percent of the given stress as under zero stress.

These observations on the recovery of the ability to creep transiently are of particular interest in interpreting the processes of creep. The fact that the units of

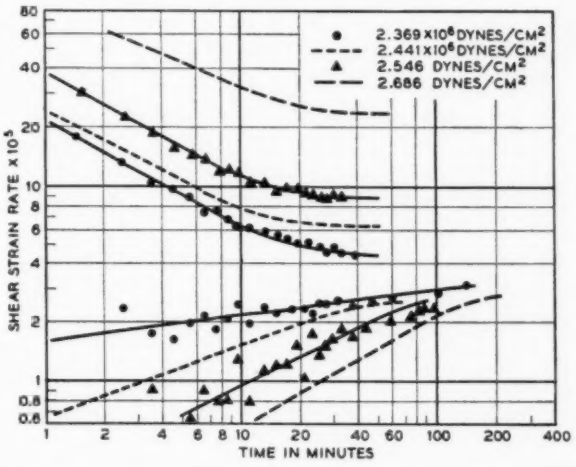


FIG. 4. Creep behavior at a resolved shear stress of 2.301×10^6 dynes/cm² after 40 min at various higher loads. The upper curves show the creep at the higher loads and the lower curves show the creep after reducing the load.

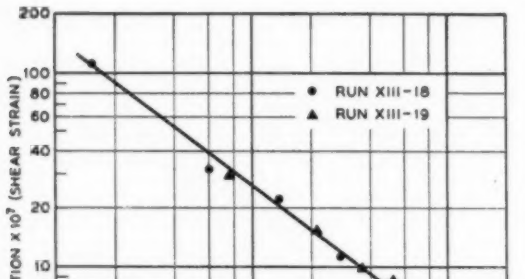


FIG. 5. Contraction of a zinc single crystal after 100 min creep at a stress of 1.96×10^6 dynes/cm².

flow which determine the creep behavior at a given stress can fully recover although the crystal is actively creeping at a stress only 5 percent lower indicates that creep at the lower stress does not affect the units which determine creep at the higher stress. In effect, this means that thermal fluctuations do not activate the same units of flow as are activated by a five percent increase in stress. The point of view taken by one of us[7] in an earlier paper, by Dushman, Dunbar, and Huthsteiner[8] and by Frederickson and Eyring[9] is, therefore, not valid as it applies to the analysis of creep data in metals.

The finding that thermal fluctuations play a comparatively small role in the activation of units of flow when compared with the role of the stress is not an entirely new one. The existence of this situation has been suspected from the fact that single crystals can creep at temperatures close to absolute zero at stresses not orders of magnitude larger than the yield points found at ordinary temperatures.[10] Nevertheless, even though thermal fluctuations are relatively ineffectual in supplying the activation requisite for creep, they seem to be fairly effective in reducing the required activation energy by altering either the units of flow or their environment. This would indicate that in order to produce plastic strain, an action which is coordinated over a fairly large volume of the crystal is required. Stress provides such an action but thermal fluctuations do not. On the other hand, by acting first at one point and then at another thermal fluctuations can progressively lower the required activation energy until the existing stress can move the unit and produce strain.

[7] W. J. Kauzmann, Trans. Am. Inst. Mining Met. Engrs. **143**, 57 (1941).
[8] Dushman, Dunbar, and Huthsteiner, J. Appl. Phys. **15**, 108 (1944).
[9] J. W. Frederickson and H. Eyring, Metals Tech. **15**, No. 5, Technical Publication No. 2423 (1948).
[10] E. Schmid and W. Boas, *Plasticity of Crystals* (F. A. Hughes, London, 1950), p. 146.

The frequent occurrence of the linear relationship between the logarithm of creep rate and the logarithm of time is also felt to be of fundamental significance in the behavior of units of flow which cause creep. A relationship of this type has been observed for: (a) The decrease in rate in transient creep. (b) The increase in rate from a value near zero after a small decrease in stress during quasi-viscous creep. (c) The decrease in rate of contraction after completely removing the stress on a plastically strained crystal.

We shall now show how this behavior might arise.

Consider a process which is characteristic of a crystal and whose rate is P. Let there be an energy barrier τ which must be surmounted by means of thermal fluctuations before the process can occur. Then

$$P=A\exp(-\tau/RT), \tag{1}$$

where A is a constant, R is the gas constant, and T is the absolute temperature.

Now consider τ to be a variable energy barrier and let there be an activation energy Q associated with a change in τ. Thus,

$$\frac{d\tau}{dt}=B\exp(-Q/RT), \tag{2}$$

where B is a constant and t is the time.

There is the possibility that Q depends on τ so that

$$Q=E\pm b\tau, \tag{3}$$

where E and b are constants. Thus we can rewrite Eq. (2) as

$$\frac{d\tau}{dt}=B\exp[-(E\pm b\tau)/RT]. \tag{4}$$

Kuhlman[11] has shown that if Eq. (4) is valid, then

$$\tau=C\pm D\ln t, \tag{5}$$

where C and D are constants. Substituting in Eq. (1), we obtain

$$\log P=n\log t+\log a, \tag{6}$$

or

$$P=at^n,$$

where a and n are constants. This is the behavior which has been mentioned in (a), (b), and (c), above. We may, therefore, conclude that the activation energies for these processes obey an equation of the form (4) and change with time.

It is easily shown that these conditions imposed on the activation energies can be explained by assuming that the units of flow are dislocations and that dislocations multiply during transient creep. (A completely analogous argument can be presented from the viewpoint of Frank-Read generators which harden during creep.) Because of the forces between dislocations, the increase in their number makes their movement in-

[11] D. Kuhlman, Z. Physik **124**, 468 (1947).

creasingly difficult thus causing an increase in the activation energy for creep and a corresponding decrease in creep rate which would be expected to be controlled by equations of the form (1) to (6).

As the dislocations multiply and interact, there is a tendency for the crystal to reduce the number of dislocations, the driving force for the removal of dislocations being primarily the increase in elastic energy associated with these imperfections. Eventually the tendency reaches a point at which the rate of disappearance of dislocations is equal to their rate of production and the steady-state quasi-viscous creep is observed. If the stress is now reduced slightly, creep practically stops because the activation energy for creep is too great to be surmounted by thermal fluctuations. The fluctuations can, however, aid in removing or altering dislocations and thus lower the effective activation energy for creep so that the creep rate gradually begins to increase. As the number of dislocations decreases, the driving force for their disappearance also decreases and the creep rate increases more slowly, again in accord with Eqs. (1) to (6).

The contraction that is observed upon completely relaxing the stress can be explained by assuming that just before removing the stress there are many dislocations within the crystal which are held at obstacles which prevent their passing through the crystal. These obstacles may be impurity atoms, lattice disorders, etc. Upon removing the external stress, many of the dislocations fly back to their points of generation, being impelled by their mutual repulsions as well as by the elastic repulsions of the obstacles. These motions of the dislocations produce the instantaneous contraction. A mechanism of this sort has been proposed by Kuhlman, Masing, and Raffelseiper[12] in connection with a theory of recovery. There may, however, well be dislocations which have become entangled with obstacles to such an extent that they are not immediately able to fly back. Their mutual repulsions provide a driving force which, aided by thermal fluctuations, may help them gradually to untangle and return to their point of generation, so producing the time-dependent contraction. This driving force is again dependent on the number of dislocations and thus decreases as their number decreases. This causes an effective rise in the activation energy and the rate of the process decreases logarithmically with time.

Thus, the concept of the dislocation is consistent with the observations reported in this paper. It is evident that more experimental observations along these lines are required to produce sufficient data to allow a quantitative test of the proposed mechanisms. It is felt, however, that the techniques of studying the transient effects of slight variations of stress during creep, as well as recovery under load, have been overlooked in the past and could be useful in elucidating the phenomena of creep.

[12] Kuhlman, Masing, and Raffelseiper, Z. Metallkunde **40**, 241 (1949).

Part II. The Hardening of Pure Zinc Crystals on Annealing at High Temperatures

Slifkin[6] has reported that some crystals of zinc show the anomalous behavior of becoming much harder after resting several hours at room temperature. He called this effect rest hardening. Orowan[13] had earlier reported a similar effect (which he called thermal hardening) while studying the stress-strain relationship in zinc crystals. Wain and Cottrell[14] have verified Orowan's results but only if the zinc is exposed to nitrogen gas in the presence of a flux.

Of all the crystals studied in this work only one, Crystal II, which was grown in an uncleaned crucible, showed the phenomenon of rest-hardening on standing at 35°C. After a few runs with this crystal, however, it, too, no longer showed the effect. We are inclined to agree with Wain and Cottrell that rest-hardening is associated with impurities and we feel that the zinc used in the present work was purer than that used by Slifkin and Kauzmann.‡

In the present series of experiments we have observed another interesting type of hardening phenomenon which takes place only at high temperatures and which does not seem to require impurities. (At least it does not seem to require as much impurity as rest-hardening.) When our zinc crystals were annealed above about 200°C they were found to be considerably harder than crystals which had recovered at room temperature. (The creep rate under a given stress was a factor of 10 smaller than the rate under the same stress if recovery had taken place at low temperatures.) This phenomenon was observed in every one of the 18 crystals which were studied. It will be referred to as "anneal-hardening."

The anneal-hardening of a typical crystal (Crystal XI) is shown in Fig. 6. The upper curve is a typical creep curve of the crystal in the "soft state." This run was preceded by a 20 hour anneal at 190°C. The lower curve is the creep of the same crystal under the same stress after annealing 3 hours at 225°C. Long annealing at 225°C or annealing up to 260°C did not appreciably alter the amount of hardening.

The annealing treatment required to produce hardening varied from crystal to crystal. For a 15 hour annealing, the temperatures required to produce observable anneal-hardening varied from 200 to 260°C in the 18 crystals studied. The effect of annealing up to 400°C was studied on one crystal and the hardening was found to be even greater after exposure to these much higher temperatures. In several crystals, attempts were made

[13] E. Orowan, Z. Physik **89**, 634 (1934); Proc. Phys. Soc. (London) **52**, 14 (1950).

[14] H. L. Wain and A. H. Cottrell, Proc. Phys. Soc. (London) **B63**, 339 (1950).

‡ A deliberate attempt was made to introduce nitrogen into Crystals XV to XVIII by exposing the molten zinc (which had previously been vacuum sublimed) to purified nitrogen for three hours with agitation. No flux was used and a film was produced which gave off ammonia on treatment with dilute hydrochloric acid. These crystals showed no signs of rest hardening.

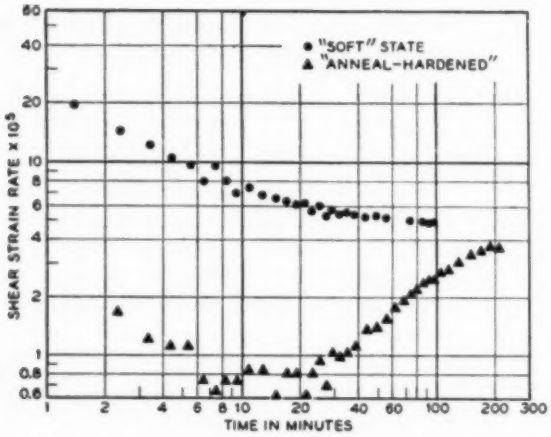


FIG. 6. Hardening of a zinc single crystal on annealing at 235°C for 3 hr (lower curve; upper curve shows normal behavior after annealing at lower temperatures).

to try to correlate the effect of time and temperature with the amount of hardening produced. Unfortunately, although the effect is always qualitatively reproducible, we were unable to reproduce quantitatively the effect of a given annealing time and temperature. On occasion, even the "critical" temperature required to produce anneal hardening in a given crystal would unaccountably change by 20 or more degrees.

In all instances the annealing was performed under approximately 0.001 atmospheres pressure. In order to insure that the effect was not caused by a surface film a few of the crystals were electropolished after the anneal-hardening treatment. The electropolish did not noticeably reduce the hardness of the crystals.

Another interesting aspect of the phenomenon is evident in Fig. 6. After creeping for approximately 15 minutes at a slow rate, the hardened crystal began to show signs of softening and the creep rate increased, finally leveling off at the rate of quasi-viscous creep normally exhibited by the crystal in "soft state." This phenomenon, which might be called "strain-softening," was exhibited by most of the crystals examined. It is also found that anneal hardening can be removed by loading the hardened crystal at a somewhat larger stress before testing at the original stress. Here again, the quantitative behavior of strain softening may vary appreciably from one run to another on the same crystal even though the qualitative aspects were reproducible. Furthermore, in some crystals strain softening occurred very slowly after several hours of creep and in some cases it did not occur at all. Even in these latter cases, though, the crystal could be returned to the soft state by annealing below the "critical" anneal-hardening temperature after letting it creep for several minutes in the hardened state. A low-temperature annealing without prior strain did not, however, remove the anneal hardening.

The phenomenon of anneal hardening appears to be somewhat similar to the hardening observed by Orowan, Wain and Cottrell, and Slifkin except that it is much more easily produced, at least in crystals of the type used in this work. Furthermore, it does not seem to require the addition of impurities. The strain-softening phenomenon is somewhat analogous to the yield point characteristic of iron-carbon alloys. In the crystals studied, however, the concentration of spectroscopically detectable impurities was less than 0.001 percent and the crystals were grown in vacuum to reduce the concentration of dissolved gases. At present the mechanisms responsible for anneal hardening and strain-softening are not evident. They would appear, however, to be fairly consistent properties of large single crystals of very pure zinc, carefully grown and measured under low stress and at low total deformation.

ACKNOWLEDGMENTS

During the course of this research, Mr. Tanenbaum was the du Pont Fellow in Chemistry at Princeton University. The authors wish to thank the E. I. du Pont de Nemours Company for this support. We also wish to express our appreciation to the New Jersey Zinc Company for the gift of the 99.999+ percent zinc.

Resistivity Striations in Germanium Crystals

PAUL R. CAMP
RCA Laboratories Division, Princeton, New Jersey
(Received September 26, 1953)*

Germanium metal, both single crystalline and polycrystalline, is frequently found to contain sharp fluctuations in impurity concentration. These are superimposed on the gradual variations expected as the result of ordinary segregation processes.

A rather simple technique for detecting fluctuations of this kind has been developed. It is more sensitive than resistivity scanning and does not involve the use of radioactive tracers. This technique has been applied to a number of samples. The origin, value, and elimination of these fluctuations are discussed briefly.

I. INTRODUCTION

LARGE local variations of impurity concentration in single crystals have been the subject of some recent interest and various procedures for detecting them have been reported. Fluctuations of this sort in silicon have been observed by Scaff and Pfann,[1] who detected them in ingots which contained both *p*- and *n*-type impurities. The fluctuations gave rise to multiple *p*,*n* boundaries which were revealed by etching. English,[2] Burton,[3] and Hall[4] have reported similar phenomena in germanium. English melted a low melting point alloy onto the surface of the germanium and then removed it with an acid etch thereby making the fluctuations visible. In independent work, Burton showed similar striations by a radioactive autograph method and found, in addition, that the resistivity and hole lifetime varied with these fluctuations. More recently Chalmers[5] has used the radioactive autograph method to detect similar effects in lead bismuth alloys.

This paper is concerned primarily with resistivity fluctuations[6] occurring in single crystal *n*-type germanium, although such variations have been observed to occur in polycrystalline and *p*-type material as well. A rather sensitive and simple method for detecting these striations is described and their origin and suppression is discussed.

The germanium used in these experiments was, except as noted, single crystal material grown by withdrawing a seed crystal from a pot of molten germanium. The plane of the impurity concentration was the plane of the growing interface. Thus the regions of high concentration and consequently low resistivity were surfaces which had once been the growing interface. The magnitude and spacing of these variations depended on the growing conditions. They were often periodic but also occurred as isolated and nonperiodic discontinuities. Variations in resistivity of as much as 2.2 to 1 have been detected in successive laminae less than 1 mm thick.

II. DETECTION OF STRIATIONS

In semiconductors, small variations in the impurity concentration along a sample, cause corresponding changes in the conductivity of the sample which may be detected by standard potential scanning methods. However, such techniques are tedious and severely limited in resolution by the dimensions of the sample, and the currents that can be passed through it. A much more sensitive and effective technique for revealing resistivity striations makes use of electroplating onto the germanium a thin film of copper, which makes the striations visible. A sample is cut from the crystal to be examined, and a lead soldered onto one end. It is then ground, polished, and immersed in an acid copper sulfate plating solution. Pulses of fairly high voltage are applied to the sample and the resulting current causes copper to plate preferentially onto the low-resistivity portions of the crystal. Maximum visual contrast may be obtained by suitably adjusting the resistivity of the electrolyte, the magnitude and duration of the pulses, and the pulse repetition rate. The purpose of pulsing is to allow diffusion to replace the ions lost in the immediate vicinity of the rapidly plating region, thereby restoring the uniformity of the resistivity of the electrolyte.

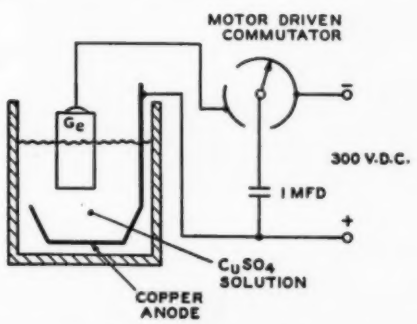


FIG. 1. Diagram of the copper plating apparatus.

* The original unrevised manuscript was submitted July 18, 1953.

[1] W. Pfann and J. Scaff, Metals Trans. **185**, 389–392 (1949).

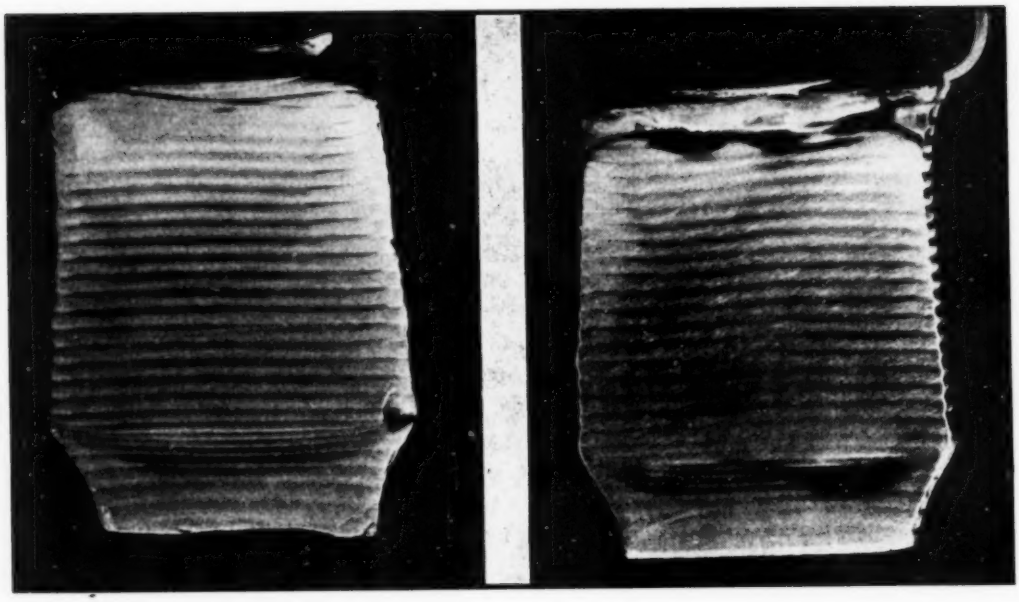
[2] A. C. English, Inst. Radio Engrs.—Am. Inst. Elec. Engrs., Conference on Electronic Devices, Durham, New Hampshire (June, 1951).

[3] J. A. Burton, Inst. Radio Engrs.—Am. Inst. Elec. Engrs., Conference on Electronic Devices, Durham, New Hampshire (June 1951).

[4] R. N. Hall, Phys. Rev. **88**, 139 (1952).

[5] Chalmers, Stewart, Thomas, Wavehope, and Winegard, Phys. Rev., **83**, 657 (1951).

[6] These fluctuations are frequently referred to as striations. They are not to be confused with the lineage structures in tin, also called striations, reported by E. Teghtsonian and B. Chalmers, Can. J. Phys. **29**, 370–381 (1951).



(a) (b)

FIG. 2. Striations as revealed by selective plating technique. The two plates shown are adjacent sections of the same 5 ohm-cm, *n*-type Ge sample. The period of the striations is about 1 mm. The section of Fig. 2a has not been annealed. That of Fig. 2b has been annealed for 63 hours at 900°C.

A system which has been found useful (Fig. 1) consists of a 1 mf condenser which is alternately charged from a 300 volt supply, and discharged through the plating solution. A motor-operated switch provides a repetition rate of from $\frac{1}{2}$ to 3 pulses per second. A condenser discharge is used because it is a convenient way of obtaining large current pulses of short duration. Peak current densities of the order of 5A/cm² are common. Plating times of from 1 to 5 min. have been found adequate for most samples. The resistivity of the electrolyte is best adjusted to the sample being analyzed. A range of 3 to 30 ohm-cm is satisfactory. If the resistivity of the solution is too high, the contrast will suffer. If it is too low, only the region near the surface of the electrolyte will become plated.

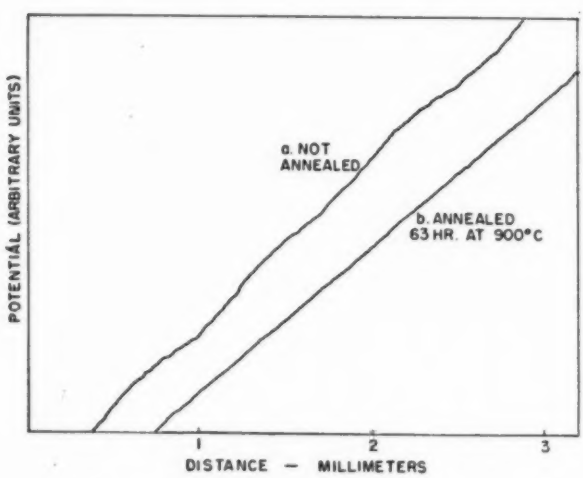


FIG. 3. Potential microscan of the two samples shown in Fig. 2.

That this method is very sensitive is illustrated by Fig. 2. The two slabs shown are cut from the same crystal of germanium (average resistivity about 5 ohm-cm). The faces shown were originally in contact. The slab on left (a) is as it was grown, the slab on the right (b) has been treated to reduce the striations as will be discussed shortly. A fine-scale potential microscan (Fig. 3), was made of these two samples and the resistivity in the left-hand crystal was found to vary by 1.8/1 from the unplated to the plated region. In the right-hand crystal, the variation was only 1.15/1 or less. This ratio is nearly the limit of resolution of the microscan for striations whose period is about 1 mm. The photograph not only shows clearly the striations in the right-hand sample but shows that the striations in the left-hand sample have fine structure.

Almost invariably, there is associated with these abrupt changes in resistivity some change in the surface structure of the crystal. Frequently, this is of the nature of slight changes in the diameter of the crystal. Figure 4 shows the surface of the crystal from which the samples of Fig. 2 were cut. The period of the surface structure is the same as that of the striations.

III. CAUSE OF STRIATIONS

The striations are believed to result from changes in the rate of advance of the growing interface. When the rate is slow, the impurities rejected by the solid are largely dispersed throughout the melt by diffusion and mixing. When the rate is fast, however, a bow wave

of impurities forms at the growing interface. This results in an artifically high concentration of impurities in the melt which the interface sees, and a consequent increase in the impurity is accepted by the solid.[7]

The rate of advance of the growing interface may be changed by a number of means. The possibilities of cyclic change are particularly interesting, since very often, the striations have a cyclic character. Such periodic changes may be artificially produced by cycling the temperature or pulling speed, or they may occur spontaneously because of some mechanical instability. One such possibility is that the meniscus is unstable; the melt being pulled by the growing crystal until a critical position is reached from which it suddenly drops back a little. This has the effect of a variable pulling speed although the external pulling speed is constant. This effect has been observed by Nelson.[8]

By far the most pronounced striations observed have resulted from slowly rotating the crystal as it was growing (Figs. 2, 5, 6). Presumably this is due to a nonuniform temperature distribution in the melt resulting in faster growth in one portion of the pot than in another. If the crystal is not rotated, this results merely in a tilting or shaping of the growing interface, which restores a constant rate to the whole interface. If the crystal is rotated, however, the result is two or more concentric spirals of material of different resistivity. Figure 5 shows the spiral structure in a sphere cut from a crystal grown under these conditions.

It was mentioned earlier that striations occur in polycrystalline material as well as in single crystals. One may capitalize on this fact and learn much about the thermal gradients within a crucible in which a melt is allowed to freeze.

Figure 6 shows the striation patterns in a polycrystalline bar. In order that good photographic reproduction might be obtained, a sample was chosen from a bar which had been rotated during growth under conditions which gave rise to pronounced striations. It was pulled from the melt in the same manner as single crystals.

The twin and grain boundaries show up as abrupt changes in the plating. Note that the striations are apparently uninterrupted by such boundaries. This indicates that all parts of the growing crystal were growing at the same rate regardless of crystal orientation. Microscopic examination of the sample showed that the striations, whose period was about 1 mm were continuous across a grain boundary to within $\pm 30\mu$.

IV. SUPPRESSION OF STRIATIONS

In order to grow striation-free crystals, it is necessary that the pulling be smooth, that the temperature distribution in the melt have the proper symmetry, and that the temperature itself be well controlled. For a glowbar type furnace and pulling speeds of about 1 mm per minute it is estimated that the temperature of the pot must be constant within $\pm\frac{1}{2}$°C.[9] If meniscus troubles are encountered, it may be desirable to superimpose on the smooth rate of pulling a high, frequency vibration (e.g. 120 cycles) to break the meniscus.

Striations introduced by rotation may be avoided in several ways. Many cold spots may be deliberately created in the furnace so as to cause a structure so fine as to be harmless. Or the melt may be vigorously agitated to destroy the effect of any cold spots and to reduce the bow wave of impurities. In addition, the speed of rotation may be increased so that the period of the striations is very small. As the period is made smaller, the amplitude is also made smaller boths

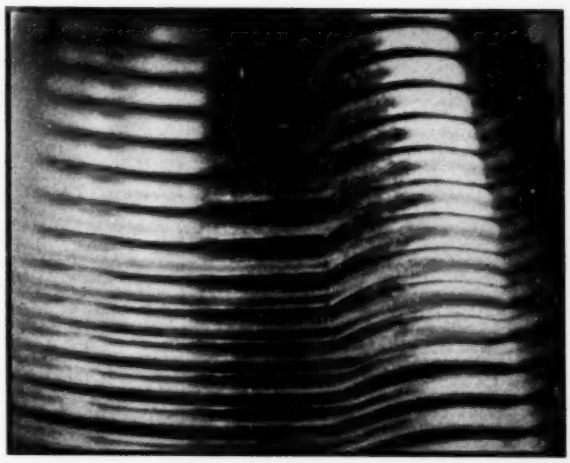


FIG. 4. Exterior of crystal from which plates of Fig. 2 were cut. Ridges have a period of about 1 mm.

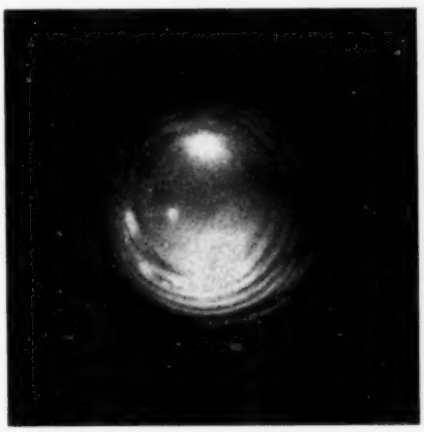


FIG. 5. Striation pattern on the surface of a germanium sphere (diam 1.8 cm) showing the spiral structure introduced by rotating the crystal during growth.

[7] The author understands that a comprehensive analysis of this process which includes the effect of circulation in the melt will be published shortly by W. Slichter and J. A. Burton.

[8] H. Nelson, RCA Harrison, private communication.

[9] This figure is cited for a glowbar type furnace with a long thermal time constant. The requirements for an rf furnace with small thermal inertia may be quite different.

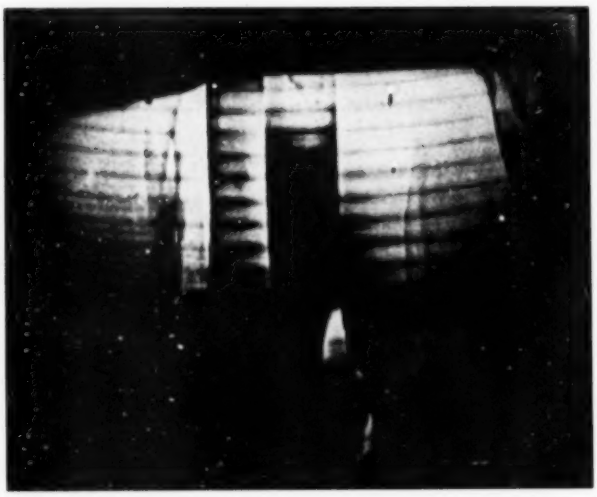


FIG. 6. Striations in a polycrystalline slab illustrating the continuity of striations across grain boundaries.

because the bow wave does not have so long to build up and because the rotation causes stirring.

A word of caution is necessary. When the speed of rotation is increased, or when the melt is agitated, stable flow patterns may be produced which result in one portion of the crystal being permanently in a region of higher flow rate than another. Thus one portion of the growing crystal (for example, the region near the center of a rotating crystal) may be confronted by a melt with a different impurity concentration than that facing another. This results not in periodic variations of resistivity along the crystal, but rather in variation across the crystal. Therefore, care must be taken to insure either very uniform or quite random circulation of the melt.

It has been found that the proper annealing treatment greatly reduces the amplitude of the resistivity striations. It should be emphasized however, that if homogeneous crystals are desired, it is better to grow unstriated crystals than to try to reduce the striations once they appear.

Figure 3 shows a potential micro-scan of the two samples of germanium of Fig. 2 which were originally badly striated. The two plates are adjacent longitudinal slabs of the same sample. The pieces are about 1.5 cm long and were probed lengthwise. Curve (a) is for one bar which was untreated, and curve (b) is for the other which had been annealed for 63 hours at 900°C. Curve (a) shows resistivity variations in adjacent zones of the order of 1.8/1. In curve (b) the resistivity variations have been reduced to 1.15 or less. Note that the fine structure of the left-hand picture of Fig. 2 is not resolved by the potential scan, and that the striations in the right-hand picture, just barely resolved by the micro-scan, stand out clearly. The fine structure has evidently been eliminated by annealing.

Although for most purposes it is desirable that crystals be free from striations, their presence is sometimes welcome. Since the striations show successive positions of the growing interface, detailed studies of the growth of crystals or the freezing of polycrystalline ingots may be made. The pulse plating technique is not only one of the most sensitive methods for doing this, but is fast, requires no special apparatus or radiation laboratory, and most important, may be used without injuring the crystals or having planned such an examination before the crystal was grown.

V. EXPLANATION OF THE SELECTIVE PLATING PHENOMENON

When this dependence of plating on resistivity was first discovered, it was assumed that the copper deposit simply mapped the current density which was controlled by the average resistance of the various current paths through the crystal. There were a number of features which supported this view. The resistivity was in the proper phase with the plating. Surfaces could be reground, polished, and replated, and still show the striation pattern. An etched surface would plate selectively. If surfaces were etched in a particular way, they showed a washboard structure which had the same period as the striations revealed by plating. In cutting the crystal, it was often noted that the saw left marks of the same period as the striations, indicating fluctuation of mechanical properties as well as electrical.

However, there have developed a number of serious objections to this explanation. The measured resistivity fluctuations can be well below 2:1 for crystals which show a variation of plating of more nearly 10:1. Slight contamination of the surface may obscure the striations. There is a very definite dependence on current density. At low-current densities (of the order of one milliampere per sq cm), the plating is uniform. As the current density is increased, the striations begin to appear. Further increase in current density increases the contrast. One of the most convincing arguments is that the striations can be shown in fine detail with sharp outlines on thick crystals in which it is known that striations bend and interweave.

Although the current density dependence strongly suggests a nonlinear process, it has been found that the current voltage characteristic of the whole Ge $CuSO_4$ cell is linear in the range used. An explanation that seems reasonable, therefore, is that there exists at the surface of the germanium an interface layer with nonlinear electrical characteristics which are directly related to the properties of the bulk germanium underneath. A sufficiently thin layer would have a small resistance compared to that of the sample and could control the current distribution through the surface without appreciably affecting the total current.

Several experiments support this hypothesis and suggest that the layer in question is an oxide film. If the sample is bathed in fairly concentrated HNO_3 before plating, the contrast is visibly improved. If it is bathed in HF (which would dissolve an oxide film) and immediately plated, the contrast is almost completely destroyed. If, however, it is bathed in HF and then washed in distilled water before plating, the contrast is largely restored. Apparently a new film is able to form during the washing.

JOURNAL OF APPLIED PHYSICS VOLUME 25, NUMBER 4 APRIL, 1954

Motion of Small Particles in Skew Shape Suspended in a Viscous Liquid

CHAN-MOU TCHEN
National Bureau of Standards, Washington, D. C.
(Received August 21, 1953)

The resistance experienced by a curved and elongated small particle is studied by the method of velocity perturbations. The fundamental equations determining the perturbations are derived from the equations of Oseen. The particle is curved in such a way that the axis of the particle is formed by an arc of circle. The opening angle may vary to exhibit various shapes from a straight ellipsoid to a closed ring. Results illustrate the circumstances under which manifests the cooperation between the different parts of the particle. Apart from this pure hydrodynamical interest, the problem has applications to long chain molecules in solution and to suspensions of swimming organisms. It demonstrates the effect of skew shape by taking for the chain molecule a model in the form of a variable-curved ellipsoid instead of a sphere or a cluster of spherical symmetry, as the coordination of the chain links, even in the case of free rotation, leads initially to configurations of variable skewness generally obscured under the averaging processes by current theories.

1. INTRODUCTION

IN the study of the motion of small particles suspended in a viscous liquid, there are two groups of problems which arise in the first place: those dealing with the influence of the presence of the particles upon the apparent or effective viscosity of the liquid, if the latter possesses a non uniform motion, and those relating to the frictional resistance of the particles moving, for example, linearly in a liquid at rest. The problems of the first group, on their part, lead to the question of the orientation of the particles, orientation which depends at the same time upon the rotating motion communicated by the liquid to the particles, and upon the Brownian motion caused by thermal agitations. Then, there are problems concerning the mutual influences between different particles, the influences from the walls, etc. The problems of the second group raise questions of the practical applicability and the extension of the Stokes law of resistance,[1] when such underlying assumptions as absence of inertial terms in the hydrodynamic equations, rigidity of the particle, unboundedness of the liquid surrounding the particle, and stationarity of the motion are not present, and when particles other than spherical ones are considered. The first assumption cited above is generally acceptable for the study of resistance of small particles in laminar motion. If the second and third assumptions are removed, for example with the case of a cluster of similar spherical particles being embedded in the liquid,[2] the modifications required in the Stokes law lead to important applications to flexible chain-molecules in solution. For this purpose, Brinkman,[3] and Debye and Bueche[4] have independently investigated a model of a molecular cluster consisting of a porous sphere, the porosity being connected with the number of links in the chain-molecule. Kirkwood and Riseman[5] have studied a cluster of spherules distributed randomly and symmetrically around a molecular center of mass. The results obtained are quite comparable with those of Debye and Bueche, as if the cluster would be substituted by a sphere. In the theoretical treatments, some of which are mentioned above, one is preoccupied almost exclusively with models of relatively symmetrical shapes: spheres or clusters with a distribution of spherical symmetry, ellipsoids,[6] and cylinders.[7] However, due to linkage, the cluster of chain-molecule possesses most probably skew configurations, like curved filaments. *Specifically, a variable porous-curved ellipsoid would seem to be a more probable configuration than a porous sphere, at least before averaging processes.*

Further, the nonstationarity in the motion of the particles and in the perturbations of the liquid opens a field which still remains obscure in the case of chain molecules in solution. Its effects on the motion of a small body may be visualized from the progression of

[1] M. S. Smoluchowski, Proc. Fifth Internatl. Cong. Math. (Cambridge) **2**, 192 (1912).
[2] J. M. Burgers, Proc. Koninkl. Ned. Akad. Wetenschap. (Amsterdam) **44**, 1045, 1177 (1941); **45**, 9, 126 (1942).
[3] H. C. Brinkman, Proc. Koninkl. Ned. Akad. Wetenschap. (Amsterdam) **50**, 618, 821 (1947); Physica **13**, 447 (1947).
[4] P. Debye and A. M. Bueche, J. Chem. Phys. **16**, 573 (1948).
[5] J. G. Kirkwood and J. Riseman, J. Chem. Phys. **16**, 565 (1948).
[6] A. Oberbeck, Crelle's Journal **81**, 62 (1876).
[7] J. M. Burgers, Chap. III of the *Second Report on Viscosity and Plasticity*, Verhandel. Koninkl. Ned. Akad. Wetenschap. Afdeel. Natuurk. (Amsterdam) (le sectie) **16**, No. 4 (1938).

small organisms by making certain unstationary tail motions or unstationary body deformations. In this connection, the problem of the motion of a curved elongated body may also play an important part.

For practical purposes, and for pure hydrodynamical interest as regarding the mutual cooperation or "shielding" in overcoming the resistance when the two arms of an elongated particle are approached, we shall investigate, in the following pages, the resistance experienced by a curved ellipsoid moving in a viscous liquid at rest. The problem of effective viscosity, Brownian motion, as well as unstationarity of the motion, will be left out of consideration.

2. CONDITIONS OF THE PROBLEM AND DESCRIPTION OF OSEEN'S METHOD

In order to fix the conditions of our problem, we suppose that the motion is very slow and that the particle is small, so that the inertial effects are negligible. This permits the linearization of the general hydrodynamic equations. We consider one single particle in an infinite liquid, so that the interactions between particles and actions of the walls are not under consideration. The particle is elongated and curved in such a way that the axis of the particle is formed by an arc of circle. The radius of curvature, and the opening angle formed by the two arms of the elongated particle are two parameters which admit to exhibit all shapes varying from a straight ellipsoid to a closed ring of uniform circular cross-section. Such a particle is assumed to move rectilinearly at constant speed in a liquid at rest. For such a stationary case, the resistance experienced by this particle will be identical in absolute value to the resistance experienced by a particle held at rest in a liquid moving at constant speed.

We start from the following equations for the velocity perturbations of the liquid produced by an external force. They are derived from the equations of Oseen.[7]

$$\left.\begin{aligned} u&=\frac{1}{8\pi\eta}\left[F_x\Delta\Psi-F_x\frac{\partial^2\Psi}{\partial x^2}-F_y\frac{\partial^2\Psi}{\partial x\partial y}-F_z\frac{\partial^2\Psi}{\partial x\partial z}\right]\\ v&=\frac{1}{8\pi\eta}\left[F_y\Delta\Psi-F_x\frac{\partial^2\Psi}{\partial x\partial y}-F_y\frac{\partial^2\Psi}{\partial y^2}-F_z\frac{\partial^2\Psi}{\partial y\partial z}\right]\\ w&=\frac{1}{8\pi\eta}\left[F_z\Delta\Psi-F_x\frac{\partial^2\Psi}{\partial x\partial z}-F_y\frac{\partial^2\Psi}{\partial y\partial z}-F_z\frac{\partial^2\Psi}{\partial z^2}\right]\end{aligned}\right\}.$$

In these equations F_x, F_y, F_z represent the components of the force per unit volume, applied at the origin of the coordinates; they are constant with respect to time. The liquid possesses an original motion, constant and uniform, of velocity U parallel to $0x$; u, v, w are the components of the velocity perturbation produced by the force at the point x, y, z situated at a distance $r=(x^2+y^2+z^2)^{\frac{1}{2}}$ from the point of application of the force; η is the coefficient of viscosity of the liquid, and Ψ represents a function of coordinates and of a "Reynolds number" given by $\rho Ur/\eta$ (ρ being the density of the liquid). For small Reynolds numbers which interest us here, this function can be written as follows:

$$\Psi=r.$$

In the case where the forces with components $f(x',y',z')ds$; $g(x',y',z')ds$; 0; are distributed continuously along the elements of a curve s, the coordinates of which being given by $x'(s)$, $y'(s)$, $z'(s)$, the velocity perturbations will be determined by the following expressions:

$$\left.\begin{aligned} 8\pi\eta u&=\int ds f\left\{\frac{1}{r}+\frac{(x-x')^2}{r^3}+\right\}\int ds g\frac{(x-x')(y-y')}{r^3}\\ 8\pi\eta v&=\int ds f\frac{(x-x')(y-y')}{r^3}+\int ds g\left\{\frac{1}{r}+\frac{(y-y')^2}{r^3}\right\}\\ 8\pi\eta w&=\int ds f\frac{(x-x')(z-z')}{r^3}+\int ds g\frac{(y-y')(z-z')}{r^3}\end{aligned}\right\}, \quad (1)$$

where the integrals are extended along the line s, the new distance being $r=[(x-x')^2+(y-y')^2+(z-z')^2]^{\frac{1}{2}}$.

Consider now an elongated particle, quasi-cylindrical, with circular section of radius b, the axis of which is curved along an arc of circle of radius a. The opening angle formed by the two arms may be left arbitrary. In the following, it will not be necessary to assume that b be a constant. One may suppose that b will depend upon the position of the section along the axis; the particle, for example, will have a maximum thickness in the middle, and from it gradually decreases toward the two ends, such as the case with a curved ellipsoid. We shall even see that the introduction of a special relation of this kind between b and the position of the considered section will help us to simplify the final results of the calculations.

We suppose first that the particle is held at rest, the center of the circle being at 0 (see Fig. 1), while the particle is symmetrical with respect to $0yz$, and we want to calculate first the resistance experienced by this particle due to the motion, say U, of the liquid.

The problem of the resistance can be studied in the following way: we replace the particle by a system of forces fds, gds, 0, distributed in a continuous manner along the axis of the particle. The velocity perturbations, produced by these forces at a point x, y, z are determined by formulas (1).

The mean values of the velocity perturbations over any normal section of the particle can then be calculated. These components will be denoted by $\langle u\rangle$, $\langle v\rangle$, 0. As can be seen immediately the component $\langle w\rangle$ vanishes. Multiplied by a frictional coefficient ζ, they represent the reactions produced by the particle upon the liquid in motion at that normal section. The total

resistance, or in other words the total reaction changed in sign, is obtained by an integration over the whole particle, and has the following components:

$$W_x=-\zeta\int ds\langle u\rangle, \tag{2a}$$

$$=2a\chi_0\zeta U, \tag{2b}$$

$$W_y=-\zeta\int ds\langle v\rangle=0. \tag{2c}$$

Here, $2a\chi_0$ is the total length of the particle.

3. MEAN PERTURBATION OVER A CROSS SECTION

For the purpose of calculating the cross-sectional mean perturbations, we shall start by considering, in the cross section defined by the angle χ (Fig. 1), a point M, (of coordinates x, y, z), determined by the angle θ. If P (of coordinates $x', y', 0$, defined by the angle χ') represents the center of one of the elements where forces fds and gds are exerting, the distance PM is given by the following relation:

$$r^2=a^2+(a+b\cos\theta)^2-2a(a+b\cos\theta)\cos(\chi-\chi)+b^2\sin^2\theta.$$

By writing

$$\chi'-\chi=2\nu, \tag{3}$$

we shall have

$$r^2=2a^2(1-\cos2\nu)+2ab\cos\theta(1-\cos2\nu)+b^2,$$

or

$$r^2=4a^2(\sin^2\nu+\epsilon^2)\left(1+\frac{\lambda b}{a}\cos\theta\right),$$

where

$$\epsilon=b/2a,$$

$$\lambda=\frac{\sin^2\nu}{\sin^2\nu+\epsilon^2}. \tag{4}$$

According to what was stated at the beginning of Sec. 3, ϵ may be a function of χ.

By writing

$$\Delta=\sin^2\nu+\epsilon^2, \tag{5}$$

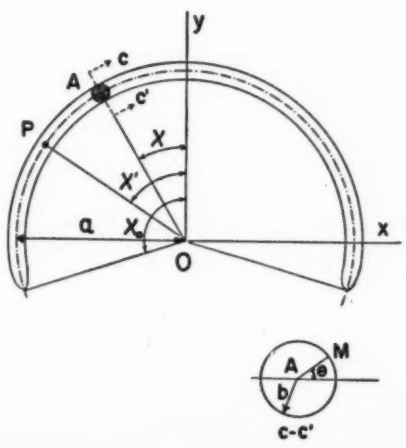


FIG. 1. Shape of the particle.

we shall have, finally,

$$r^2=4a^2\Delta(1+2\lambda\epsilon\cos\theta).$$

As we suppose that b is small compared with a, such that $\epsilon\ll1$, and as $\lambda<1$, we can write by approximation

$$\frac{1}{r}=\frac{1}{2a\Delta^{\frac{1}{2}}}(1-\lambda\epsilon\cos\theta),$$

$$\frac{1}{r^3}=\frac{1}{8a^3\Delta^{\frac{3}{2}}}(1-3\lambda\epsilon\cos\theta),$$

where we have neglected the terms of order ϵ^2 as compared with unity. It is not necessary to notice that the term ϵ^2 occurring in the expression (5) is not negligible, since the term which precedes it is variable and can even become zero.

The components of PM along the axes of coordinates are:

$$\left.\begin{aligned}x-x'&=-(a+b\cos\theta)\sin\chi+a\sin\chi'\\ y-y'&=(a+b\cos\theta)\cos\chi-a\cos\chi'\\ z-z'&=b\sin\theta\end{aligned}\right\}. \tag{6}$$

These values must be brought now to Eq. (1). Having done this we must calculate the mean value of the fractions with respect to θ. We see immediately that the expression of w, which involves a factor $\sin\theta$, vanishes in the average, as can be justified by the symmetry. Consequently, we need to consider only the first two expressions of (6), which we write in the following form:

$$x-x'=a(\sin\chi'-\sin\chi)-b\sin\chi\cos\theta,$$

$$y-y'=-a(\cos\chi'-\cos\chi)+b\cos\chi\cos\theta.$$

These expressions will be useful to calculate the mean value of the fractions as occurring in (1). To abbreviate the formulas, we shall write:

$$\sin\nu=S;\quad\cos\nu=C.$$

By observing that

$$ds=ad\chi'=2ad\nu$$

we find, after some transformations:

$$\left.\begin{aligned}\left\langle\frac{1}{r}\right\rangle ds&=\Delta^{-\frac{1}{2}}d\nu\\ \left\langle\frac{(x-x')^2}{r^3}\right\rangle ds&=\tfrac{1}{2}\Delta^{-\frac{1}{2}}[(\epsilon^2/2+S^2)\\ &\quad-(\epsilon^2/2-S^2+2S^4)\cos2\chi-2S^3C\sin2\chi]d\nu\\ \left\langle\frac{(y-y')^2}{r^3}\right\rangle ds&=\tfrac{1}{2}\Delta^{-\frac{1}{2}}[(\epsilon^2/2+S^2)\\ &\quad+(\epsilon^2/2-S^2+2S^4)\cos2\chi+2S^3C\sin2\chi]d\nu\\ \left\langle\frac{(x-x')(y-y')}{r^3}\right\rangle ds&=\tfrac{1}{2}\Delta^{-\frac{1}{2}}[-(\epsilon^2/2-S^2+2S^4)\\ &\quad\times\sin2\chi+2S^3C\cos2\chi]d\nu\end{aligned}\right\}. \tag{7}$$

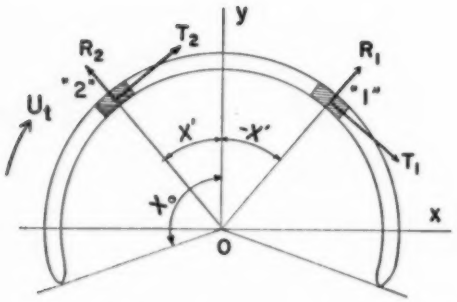


FIG. 2. Symmetry and antisymmetry of forces distributed along the axis of the particle.

These expressions brought to (2) will give the mean value of the velocity perturbations over the cross section in terms of the forces f and g.

4. CALCULATIONS OF RESISTANCE IN THE $0x$ AND $0y$ DIRECTIONS

According to Eqs. (1), the mean resultant velocity of components $U+\langle u\rangle$, $\langle v\rangle$, 0, at a cross section defined by the angle χ, are determined from the forces $f(\chi')$ and $g(\chi')$, through the fractional expressions. The forces are *a priori* unknown, as well as the resultant velocity. If a velocity is written proportional to a force in our case of laminar motion, then the Eqs. (1) can be considered as a system of integral equations determining the forces with the fractional expressions playing the role of kernels.

With the values given by (7), the kernels are functions of χ and χ' entering separately in a complicated way. It is therefore desirable to look for transformations which will simplify the forms of the kernels, and especially in such a way that the two angles χ and χ' enters mainly as a difference. This is made possible by changing the variables f, g, $\langle u\rangle$, $\langle v\rangle$, respectively, into T, R, u', v' as follows:

$$\left.\begin{aligned} T(\chi') &= f(\chi')\cos\chi' + g(\chi')\sin\chi' \\ R(\chi') &= -f(\chi')\sin\chi' + g(\chi')\cos\chi' \\ u'(\chi) &= \langle u\rangle\cos\chi + \langle v\rangle\sin\chi \\ v'(\chi) &= -\langle u\rangle\sin\chi + \langle v\rangle\cos\chi \end{aligned}\right\}. \tag{8}$$

It is immediately seen that the new variables T, R, are, respectively, the tangential and radial components of the force, and u', v' are the tangential and radial components of the velocity perturbation. In terms of the new variables (8), the equations of mean perturbations as derived from (1) will become much simpler, and after some transformations can be written as follows:

$$\left.\begin{aligned} u' &= \frac{1}{8\pi\eta}\int_{-(\chi_0/2)-(\chi/2)}^{(\chi_0/2)-(\chi/2)} d\nu\,[T(\chi')L(\nu)+R(\chi')M(\nu)] \\ v' &= \frac{1}{8\pi\eta}\int_{-(\chi_0/2)-(\chi/2)}^{(\chi_0/2)-(\chi/2)} d\nu\,[T(\chi')M'(\nu)+R(\chi')L'(\nu)] \end{aligned}\right\} \tag{9}$$

with

$$\left.\begin{aligned} L &= \Delta^{-1}(\epsilon^2+2S^2-3S^4) \\ M &= -2SC\Delta^{-1}(\epsilon^2+\tfrac{3}{2}S^2) \\ L' &= \Delta^{-1}(\tfrac{3}{2}\epsilon^2+S^2-3S^4) \\ M' &= \tfrac{3}{2}\cdot 2SC\Delta^{-1}(\epsilon^2+S^2) \end{aligned}\right\}. \tag{10}$$

Here, 2ν is $\chi'-\chi$, as given by (3), ϵ is defined by (4), Δ is defined by (5), S and C are abbreviated symbols for $\sin\nu$ and $\cos\nu$, as used earlier in (7). The integrations in (9) extend over the whole length of the particle. Finally in these notations, the expressions (2) for the total resistance become:

$$\left.\begin{aligned} &\text{(a) tangential} \quad W_t = -\zeta a\int_{-\chi_0}^{\chi_0} d\chi\, u' = 2a\chi_0\zeta U_t \\ &\text{(b) radial} \quad\quad W_r = -\zeta a\int_{-\chi_0}^{\chi_0} d\chi\, v' \end{aligned}\right\}. \tag{11}$$

Now we shall first calculate the resistance W_t experienced by the particle suspended in a liquid moving tangentially to the particle with a constant speed U_t. For this purpose we shall calculate u', v' from (9). The kernels L, M, L', M' are still too complicated, so that the system of integral equations (9) cannot be solved rigorously.

Before going to the calculations, it is advantageous to note a property of symmetry to be satisfied by the functions T and R. Consider two elements "1" and "2" situated symmetrically with respect to $0y$, and having positions defined by the angles $-\chi'$ and χ' (see Fig. 2). On those elements exerted the forces:

$$\text{On "1"}\begin{cases} T_1 a d\chi' \text{ directed clockwise} \\ R_1 a d\chi' \text{ directed outward} \end{cases}$$

$$\text{On "2"}\begin{cases} T_2 a d\chi' \text{ directed clockwise} \\ R_2 a d\chi' \text{ directed outward} \end{cases}$$

As the equations which serve to determine T and R are linear in T, R, and U_t, it is evident that by changing the sign of U_t—i.e., by changing the direction of motion of the liquid—, the values of T_1, R_1, T_2, R_2 will also change their signs. Moreover, the situation obtained must be the same as that resulted from a reflection of the original situation with respect to the plan $0yz$. It follows that we must have

$$T_1 = T_2,$$
$$R_1 = -R_2.$$

We may say that T is an even function and R an odd function of χ'. Supposing that those functions can be represented by Fourier series, the development can

be written, by limiting to the terms in $2\chi'$:

$$\left.\begin{aligned} T&=\alpha_0+\alpha_1\cos\chi'+\alpha_2\cos 2\chi' \\ R&=\beta_1\sin\chi'+\beta_2\sin 2\chi' \end{aligned}\right\}. \tag{12a}$$

By substituting $\chi'=\chi+2\nu$, we can write

$$\left.\begin{aligned} T&=T_0+T_2S^2+T_4S^4+2T_0^*SC+2T_2^*S^3C \\ R&=R_0+R_2S^2+R_4S^4+2R_0^*SC+2R_2^*S^3C \end{aligned}\right\}, \tag{12b}$$

with

$$\begin{aligned} T_0&=\alpha_0+\alpha_1\cos\chi+\alpha_2\cos 2\chi, \\ T_2&=-2\alpha_1\cos\chi-8\alpha_2\cos 2\chi, \\ T_4&=8\alpha_2\cos 2\chi, \\ T_0^*&=-\alpha_1\sin\chi-2\alpha_2\sin 2\chi, \\ T_2^*&=4\alpha_2\sin 2\chi, \\ R_0&=\beta_1\sin\chi+\beta_2\sin 2\chi, \\ R_2&=2\beta_1\sin\chi-8\beta_2\sin 2\chi, \\ R_4&=8\beta_2\sin 2\chi, \\ R_0^*&=\beta_1\cos\chi+2\beta_2\cos 2\chi, \\ R_2^*&=4\beta_2\cos 2\chi. \end{aligned}$$

Here, α, β are coefficients to be determined later by identifications, and $S=\sin\nu$, $C=\cos\nu$. Having the expressions of T and R given by (12b), the integrals of (9) can be calculated, with some artifices which we shall indicate in Appendix I. As the results of integrations, (9) yields:

$$\begin{aligned} 8\pi\eta u' &= \left(2\sigma-8+6\cos\frac{\chi_0}{2}\cos\frac{\chi}{2}\right)\alpha_0 \\ &+\cos\frac{\chi}{2}\cdot\left\{\left(3\cos\chi_0+\cos\frac{3\chi_0}{2}\right)(\alpha_1+\beta_1)\right. \\ &+5\cos\frac{\chi_0}{2}\cdot\alpha_1+\left(\cos\frac{3\chi_0}{2}+\frac{3}{5}\cos\frac{5\chi_0}{2}\right)(\alpha_2+\beta_2) \\ &\left.+\frac{2}{3}\cos\frac{3\chi_0}{2}\cdot\alpha_2\right\} \\ &+\cos\chi\cdot\{2(\sigma-8)\alpha_1-4\beta_1\} \\ &+\cos\frac{3\chi}{2}\cdot\left\{3\cos\frac{\chi_0}{2}\cdot\alpha_1+11\cos\frac{\chi_0}{2}\cdot\alpha_2\right\} \\ &+\cos 2\chi\cdot\left\{\left(2\sigma-\frac{32}{3}\right)\alpha_2-\frac{8}{5}(\alpha_2+\beta_2)\right\} \\ &+\cos\frac{5\chi}{2}\cdot 3\cos\frac{\chi_0}{2}\cdot\alpha_2, \end{aligned} \tag{13a}$$

$$\begin{aligned} 8\pi\eta v' &= \sin\frac{\chi}{2}\cdot\left[-6\cos\frac{\chi_0}{2}\cdot\alpha_0\right. \\ &+\left(3\cos\frac{\chi_0}{2}-\cos\frac{3\chi_0}{2}\right)(\alpha_1+\beta_1)-2\cos\frac{\chi_0}{2}\cdot\beta_1 \\ &\left.+\left(\cos\frac{3\chi_0}{2}-\frac{3}{5}\cos\frac{5\chi_0}{2}\right)(\alpha_2+\beta_2)-\frac{2}{3}\cos\frac{3\chi_0}{2}\cdot\beta_2\right] \\ &+\sin\chi\cdot[(\sigma+3)\beta_1-4(\alpha_1+\beta_1)] \\ &+\sin\frac{3\chi}{2}\cdot 6\cos\frac{\chi_0}{2}\cdot\beta_2 \\ &+\sin 2\chi\cdot\left[\left(\sigma-\frac{7}{3}\right)\beta_2-\frac{8}{5}(\alpha_2+\beta_2)\right] \\ &+\sin\frac{5\chi}{2}\cdot 2\cos\frac{\chi_0}{2}\cdot\beta_2. \end{aligned} \tag{13b}$$

In those expressions, σ represents the value of an integral as calculated in Appendix I.

The expressions (13a) and (13b) must now satisfy as much as possible the conditions

$$\begin{aligned} u'+U_t&=0, \\ v'&=0, \end{aligned} \tag{14}$$

for values of χ between $-\chi_0$ and $+\chi_0$. By analyzing those conditions, we must notice that the angle χ enters also in the expression of σ; thus it is necessary to introduce the explicit expression

$$\sigma=\ln\left(\frac{16}{\epsilon^2}\tan\frac{\chi_0+\chi}{4}\tan\frac{\chi_0-\chi}{4}\right), \tag{15}$$

where, according to what has been remarked before, ϵ may also be a function of χ. However, we can obtain a notable simplification of our problem, if we introduce the supposition that the dependence of ϵ as function of χ be such that σ takes a *constant* value, independent of χ. In connection with the constant value of σ and the resulting profile of the particle, we shall discuss in detail in Sec. 6. For the moment, we suppose that σ is independent of χ, and return to the conditions (14). In order to satisfy those conditions, it should be necessary to give to the unknown coefficients α, β, such values as the coefficients of the goniometrical functions of χ in (13) would vanish. But it is evident that the number of unknowns is insufficient to cancel all goniometrical functions. In the first place, it seems that we must have the coefficients of

$$\cos\frac{5\chi}{2},\quad \sin\frac{5\chi}{2},\quad \cos\frac{3\chi}{2},\quad \sin\chi$$

vanished. The equations are satisfied if

$$\alpha_1=\alpha_2=\beta_1=\beta_2=0, \tag{16}$$

or in other words

$$T=\alpha_0, \quad R=0. \tag{12c}$$

Hence,

$$8\pi\eta u'=\left(2\sigma-8+6\cos\frac{\chi_0}{2}\cos\frac{\chi}{2}\right)\alpha_0, \tag{13c}$$

$$8\pi\eta v'=-6\cos\frac{\alpha_0}{2}\sin\frac{\chi}{2}\cdot\alpha_0. \tag{13d}$$

The conditions (14) are rigorously satisfied for $\chi_0=\pi$ and $\chi_0=0$. The case of $\chi_0=\pi$ corresponds to a particle in the form of a closed ring, and the case of $\chi_0=0$, and *a fortiori* $\chi=0$, corresponds to a straight ellipsoid, as will be shown in Sec. 6. For other values of χ_0, the conditions are satisfied rigorously only at $\chi=0$ (middle of the particle), and otherwise only in average values, which in the notations of (11) are:

$$a\int_{-\chi_0}^{\chi_0}d\chi u'=-2a\chi_0 U_t \tag{17a}$$

$$=\frac{2a\chi_0}{8\pi\eta}\left(2\sigma-8+\frac{6\sin\chi_0}{\chi_0}\right)\alpha_0 \tag{17b}$$

and

$$a\int_{-\chi_0}^{\chi_0}d\chi v'=0. \tag{17c}$$

Taking the average fulfillment of conditions (14) as a first approximation, and comparing (17a) and (17b), we have:

$$\alpha_0=-\frac{8\pi\eta U_t}{2\sigma-8+(6\sin\chi_0)/\chi_0}. \tag{18a}$$

It follows from (11)

$$\zeta=\frac{8\pi\eta}{2\sigma-8+(6\sin\chi_0)/\chi_0}, \tag{18b}$$

$$W_t=\frac{16\pi\eta a\chi_0 U_t}{2\sigma-8+(6\sin\chi_0)/\chi_0}, \tag{18c}$$

$$W_r=0. \tag{18d}$$

The value of α_0, α_1, α_2, β_1, β_2, obtained in (16) and (18a) seems to be the closest solution to be obtained by our method. Essentially it is equivalent to presume a *constant* uniform tangential forces, as formulated by (12c), a result analogous to that obtained for the case of a straight ellipsoid in longitudinal flow.

Higher order approximations of the problem toward the fulfillment of conditions (14) for all values of χ between $-\chi_0$ and $+\chi_0$ would require probably additional higher modes terms in the Fourier series (12a), or an assumption of orthogonal series of a more complicate type. Not only may they render the integrations impossible, but also, because of probably higher modes involved, they may play a minor role in the total resistance. For this reason we shall cease to elucidate other forms of force distributions, and consider (12c) as a reasonable basis for the approximate calculation of the total resistance.

Next we consider a fixed particle suspended in a liquid moving with a radial velocity $-U_r$. The total resistance for such a radial flow is given by the expressions:

tangential component

$$W_t=-\zeta' a\int_{-\chi_0}^{\chi_0}d\chi u'$$

radial component

$$W_r=-\zeta' a\int_{-\chi_0}^{\chi_0}d\chi v'=2a\chi_0\zeta U_r \tag{19}$$

analogous to (11). Here, ζ' is the frictional coefficient for the radial flow.

The force distribution must be symmetrical with respect to the axis $0y$. If the forces for the present case are distinguished by an asterisk, we shall have

$$f^*(\chi')=-f^*(-\chi'),$$
$$g^*(\chi')=+g^*(-\chi').$$

Then the development can be written as follows:

$$f^*=\alpha_1^*\sin\chi'+\alpha_2^*\sin 2\chi',$$
$$g^*=\beta_0^*+\beta_1^*\cos\chi'+\beta_2^*\cos 2\chi'.$$

The velocity perturbations of components u', v' can be calculated by means of formulas (9) and finally the resistance by means of formulas (19), with the same method followed for the tangential motion calculated previously. By omitting all auxiliary calculations we obtain, finally,

$$\begin{aligned}8\pi\eta u'=&\sin\frac{\chi}{2}\cdot\left[\left(5\cos\frac{\chi_0}{2}+\cos\frac{3\chi_0}{2}\right)(\alpha_1'+\beta_1')\right.\\&-11\cos\frac{\chi_0}{2}\cdot\beta_1'+\left(\frac{5}{3}\cos\frac{3\chi_0}{2}+\frac{3}{5}\cos\frac{5\chi_0}{2}\right)(\alpha_2'+)\beta_2'\\&\left.-\frac{8}{3}\cos\frac{3\chi_0}{2}\cdot\beta_2'+6\cos\frac{\chi_0}{2}\cdot\beta_0'\right]\\&+\sin\chi\cdot[(2\sigma-8)\alpha_1'+4\beta_1']\\&+\sin\frac{3\chi}{2}\cdot(8\alpha_2'-3\beta_2')\cos\frac{\chi_0}{2}\\&+\sin 2\chi\cdot\left[\left(2\sigma-\frac{184}{15}\right)\alpha_2'+\frac{8}{5}\beta_2'\right]\\&+\sin\frac{5\chi}{2}\cdot 3\cos\frac{\chi_0}{2}\cdot\beta_2',\end{aligned} \tag{20a}$$

$$
\begin{aligned}
8\pi\eta v' = {} & \left(\sigma-5+6\cos\frac{\chi_0}{2}\cos\frac{\chi}{2}\right)\beta_0' \\
& +\cos\frac{\chi}{2}\cdot\left[-\left(3\cos\frac{\chi_0}{2}+\cos\frac{3\chi_0}{2}\right)(\alpha_1'-\beta_1')\right. \\
& -2\cos\frac{\chi_0}{2}\cdot\beta_1'-\left(\cos\frac{3\chi_0}{2}+\frac{3}{5}\cos\frac{5\chi_0}{2}\right)(\alpha_2'-\beta_2') \\
& \left.-\frac{2}{3}\cos\frac{3\chi_0}{2}\cdot\beta_2'\right] \\
& +\cos\chi\cdot[4\alpha_1'+(\sigma-1)\beta_1'] \\
& +\cos\frac{3\chi}{2}\cdot 4\cos\frac{\chi_0}{2}\cdot\beta_2' \\
& +\cos 2\chi\cdot[8\alpha_2'+(\sigma+11)\beta_2'] \\
& +\cos\frac{5\chi}{2}\cdot 2\cos\frac{\chi_0}{2}\cdot\beta_2'. \qquad (20b)
\end{aligned}
$$

The approximate method, as used already before, to determine the unknown coefficients α', β', consists of satisfying as much as possible the following conditions:

$$
\begin{aligned}
u' &= 0, \\
v'+U_r &= 0.
\end{aligned} \qquad (21)
$$

The closest solutions which we obtain by our method are

$$\alpha_1'=\beta_1'=\alpha_2'=\beta_2'=0.$$

Herewith (20a) and (20b) become

$$
\left.
\begin{aligned}
8\pi\eta u' &= 6\beta_0'\sin\frac{\chi}{2}\sin\frac{\chi_0}{2} \\
8\pi\eta v' &= \left(\sigma-5+6\cos\frac{\chi_0}{2}\cos\frac{\chi}{2}\right)\beta_0'
\end{aligned}
\right\}. \qquad (22)
$$

Here again the conditions (21) are rigorously satisfied for $\chi_0=\pi$ (a closed ring) and $\chi_0\to 0$ (a straight ellipsoid). But for other values of χ_0, the conditions are satisfied rigorously only at $\chi=0$ (middle of the particle), and otherwise only in average, or in the notations of (19):

$$a\int_{-\chi_0}^{\chi_0} d\chi\, u' = 0, \qquad (23)$$

$$a\int_{-\chi_0}^{\chi_0} d\chi\, v' = -2a\chi_0 U_r \qquad (24a)$$

$$= \frac{2a\chi_0}{8\pi\eta}\left(\sigma-5+6\frac{\sin\chi_0}{\chi_0}\right)\beta_0'. \qquad (24b)$$

Comparison between (24a) and (24b) leads to

$$\beta_0' = -\frac{8\pi\eta U_r}{\sigma-5+(6\sin\chi_0)/\chi_0}. \qquad (25a)$$

It follows from (19)

$$\zeta' = 8\pi\eta\Big/\left[\sigma-5+6\frac{\sin\chi_0}{\chi_0}\right], \qquad (25b)$$

$$W_t = 0, \qquad (25c)$$

$$W_r = 16\pi\eta U_r a\chi_0\Big/\left[\sigma-5+6\frac{\sin\chi_0}{\chi_0}\right]. \qquad (25d)$$

Finally, by starting from the results (18) and (25) for tangential and radial motions, it is an easy matter to calculate the horizontal and vertical components of the resistance W_x, W_y corresponding to a motion of components U_x, U_y. The components U_x, U_y at a section defined by χ, can be expressed in terms of U_t and U_r (tangential and radial components) in an analogous way as in (8). We have, for the velocity components,

$$
\begin{aligned}
U_t &= U_x\cos\chi+U_y\sin\chi, \\
U_r &= -U_x\sin\chi+U_y\cos\chi,
\end{aligned}
$$

and, for the force components,

$$
\begin{aligned}
f(\chi) &= T\cos\chi-R\sin\chi, \\
g(\chi) &= T\sin\chi+R\cos\chi.
\end{aligned}
$$

As the forces and the velocities are proportional to each other, with frictional coefficients as factors of proportionality

$$T=-\zeta U_t, \quad R=-\zeta' U_r,$$

we can write f, g also in terms of U_t, U_r, and finally in terms of U_x, U_y as follows:

$$
\begin{aligned}
-f &= \tfrac{1}{2}U_x[(\zeta'+\zeta)-(\zeta'-\zeta)\cos 2\chi]-\tfrac{1}{2}U_y(\zeta'-\zeta)\sin 2\chi, \\
-g &= -\tfrac{1}{2}U_x(\zeta'-\zeta)\sin 2\chi+\tfrac{1}{2}U_y[(\zeta'+\zeta)+(\zeta'-\zeta)\cos 2\chi].
\end{aligned}
$$

Hence, an integration with respect to χ, over the whole length of the particle, will give the total resistance as follows:

$$
\left.
\begin{aligned}
W_x &= -a\int_{-\chi_0}^{\chi_0} d\chi\, f(\chi) \\
&= U_x a\chi_0\left[(\zeta'+\zeta)-(\zeta'-\zeta)\frac{\sin 2\chi_0}{2\chi_0}\right] \\
W_y &= -a\int_{-\chi_0}^{\chi_0} d\chi\, g(\chi) \\
&= U_y a\chi_0\left[(\zeta'+\zeta)+(\zeta'-\zeta)\frac{\sin 2\chi_0}{2\chi_0}\right]
\end{aligned}
\right\}. \qquad (26)
$$

From formulas (26) it is not difficult to calculate W_x and W_y by substituting for ζ and ζ' from (18) and (25).

5. RESISTANCE OF THE PARTICLE IN A MOTION PARALLEL TO THE DIRECTION 0z

In the case of a motion of the liquid in the direction of the axis 0z, it is evident that the distribution of forces must be symmetrical with respect to the plane $y0z$. As, on the other side, the forces must change their sign at the same time as the velocity U_z of the flow, it follows that the forces f, g must be zero, and that we have to introduce only such forces $had\chi'$ parallel to the axis 0z as represented by the expression:

$$h=C_0+C_1\cos\chi'+C_2\cos 2\chi'. \tag{27}$$

The velocity perturbations which are deduced from the Oseen equations of Sec. 2 are given by the formulas:

$$\left.\begin{aligned} 8\pi\eta u &= \int ds\, h\,\frac{(x-x')(z-z')}{r^3}\\ 8\pi\eta v &= \int ds\, h\,\frac{(y-y')(z-z')}{r^3}\\ 8\pi\eta w &= \int ds\, h\left\{\frac{1}{r}+\frac{(z-z')^2}{r^3}\right\} \end{aligned}\right\}, \tag{28}$$

where the integrations have to be extended to the whole length of the particle.

By replacing $(x-x')$, $(y-y')$, $(z-z')$ by their values given by (6) and by taking the mean value with respect to the angle θ of the fractions entering in (28), we find that the mean values of u and v disappear, while for $\langle w\rangle$, we find:

$$8\pi\eta\langle w\rangle=\int_{-(\chi_0/2)-(\chi/2)}^{(\chi_0/2)-(\chi/2)} d\nu\, h\left[\frac{1}{\Delta^1}+\frac{\epsilon^2}{2\Delta^3}\right]. \tag{29}$$

Now we write expression (27) of the function h under the following form:

$$h=h_0+h_2S^2+h_4S^4+2h_0^*SC+2h_2^*S^3C,$$

where

$$S=\sin\nu,\quad C=\cos\nu,\quad \Delta=\sin^2\nu+\epsilon^2,$$

and

$$\left.\begin{aligned} h_0&=C_0+C_1\cos\chi+C_2\cos 2\chi,\\ h_2&=-2C_1\cos\chi-8C_2\cos 2\chi,\\ h_4&=+8C_2\cos 2\chi,\\ h_0^*&=-C_1\sin\chi-2C_2\sin 2\chi,\\ h_2^*&=+4C_2\sin 2\chi., \end{aligned}\right\} \tag{30}$$

The integration of (29) can then be done, and gives:

$$8\pi\eta\langle w\rangle=h_0(F_0+\tfrac{1}{2}\epsilon^2G_0)+h_2F_2+h_4F_4+h_0I_0+h_2I_2,$$

where F_0, G_0, $\cdots I_2$ are calculated according to the method indicated by Appendix I.

After substitution of these values and those of $h_0,\cdots h_2^*$, and transformation of the products of goniometric functions into sines and cosines of the sum or difference of the arguments, we obtain, finally,

$$\left.\begin{aligned} 8\pi\eta\langle w\rangle=(\sigma+1)C_0&+4(C_1-C_2/3)\cos\frac{\chi_0}{2}\cos\frac{\chi}{2}\\ &+(\sigma-3)C_1\cos\chi+4C_2\cos\frac{3\chi_0}{2}\cos\frac{3\chi}{2}\\ &+\left(\sigma-\frac{13}{3}\right)C_2\cos 2\chi \end{aligned}\right\}. \tag{31}$$

The expression (31) must satisfy the condition:

$$\langle w\rangle+U_z=0 \tag{32}$$

for the values of χ from $-\chi_0$ to $+\chi_0$. If again σ is considered as a quantity independent of χ, we can satisfy rigorously the condition (32) by taking:

$$C_0=-\frac{8\pi\eta U_z}{\sigma+1};\quad C_1=C_2=0. \tag{33}$$

Substituting the values (33) into (30), the resistance of the particle is obtained,

$$W_z=-\int_{-\chi_0}^{\chi_0}d\chi' ah=\frac{16\pi\eta U_z a\chi_0}{\sigma+1}, \tag{34a}$$

and the frictional coefficient is

$$\zeta_z=\frac{8\pi\eta}{\sigma+1}. \tag{34b}$$

6. CONCLUSION

As the value of the definite integral (F_0) (see Appendix I), we have obtained

$$(F_0)=\ln\left\{\frac{16}{\epsilon^2}\tan\frac{\chi_0+\chi}{4}\tan\frac{\chi_0-\chi}{4}\right\}, \tag{35}$$

and call this quantity σ. As already mentioned, we can obtain a notable simplification of the problem if we introduce the assumption that the dependence of ϵ as function of χ be such that σ takes a *constant* value, independent of χ. This requires that we should have

$$\left(\frac{\epsilon}{\epsilon_0}\right)^2=\tan\frac{\chi_0+\chi}{4}\tan\frac{\chi_0-\chi}{4}\Big/\tan^2\frac{\chi_0}{4}, \tag{36}$$

where ϵ_0 is the maximum value of ϵ which is obtained for $\chi=0$.

It is to be remarked that by writing $\chi_0=\pi$ in (36), ϵ becomes independent of χ, and is equal to ϵ_0 at all cross sections. Thus we obtain a closed ring of uniform thickness. On the other hand by increasing indefinitely the radius, we obtain a straight ellipsoid of revolution, as to be seen immediately from (36) by writing

$$a\chi_0=l,\quad a\chi=x,$$

l being the half-length of the particle and x being the abscissa. By increasing a or decreasing χ_0 indefinitely, Eq. (36) reduces to

$$(\epsilon/\epsilon_0)^2+(x/l)^2=1,$$

i.e., the equation of a straight ellipsoid of revolution. Thus by varying χ_0 from 0 to π, we can obtain all shapes from a straight ellipsoid to a curved one, and finally to a closed ring of uniform thickness. For all those shapes, as defined by (36), the value of σ remains then independent of χ and is

$$\sigma=\ln\left(\frac{16}{\epsilon_0^2}\tan^2\frac{\chi_0}{4}\right), \tag{37}$$

as to be obtained by writing $\chi=0$ in (35), or by substituting (36) into (35).

In order to facilitate the comparison between the results obtained here and the available ones, we shall write l in the place of $a\chi_0$ in the expression for ϵ_0 of (37). We have

$$\frac{\sigma}{2}=\ln\frac{l}{b_0}+\ln\frac{2\tan(\chi_0/4)}{\chi_0/4}. \tag{38}$$

With these notations we shall list the formulas of resistances (18), (25), (26), and (34).

It is to be remarked that the resistance can be written as

$$\text{resistance}=2a\chi_0U\cdot\text{frictional coefficient},$$

and the frictional coefficient in its turn as

$$\text{frictional coefficient}=\frac{m\pi\eta}{\ln(l/b_0)+e}, \tag{39}$$

where the parameters m, e may depend upon χ_0 (compare Figs. 3 and 4). For example:

(a) In tengential motion, we have, by comparing (18),

$$\left.\begin{aligned}&\zeta=\frac{2\pi\eta}{\ln l/b_0+e_t}\\ \text{with}\quad&e_t=\ln\frac{2\tan(\chi_0/4)}{\chi_0/4}+\frac{3}{2}\frac{\sin\chi_0}{\chi_0}-2\end{aligned}\right\}. \tag{40a}$$

(b) In radial motion, we have, by comparing (25),

$$\left.\begin{aligned}&\zeta'=\frac{4\pi\eta}{\ln l/b_0+e_r}\\ \text{with}\quad&e_r=\ln\frac{2\tan(\chi_0/4)}{\chi_0/4}+3\frac{\sin\chi_0}{\chi_0}-\frac{5}{2}\end{aligned}\right\}. \tag{40b}$$

(c) In U_x, U_y motions, we have, by comparing (26),

$$\zeta_x=\frac{1}{2}\left[(\zeta'+\zeta)-(\zeta'-\zeta)\frac{\sin2\chi_0}{2\chi_0}\right]$$

$$\zeta_y=\frac{1}{2}\left[(\zeta'+\zeta)+(\zeta'-\zeta)\frac{\sin2\chi_0}{2\chi_0}\right],$$

where ζ and ζ' are given by (40a) and (40b), respectively. If the particle is thin such that e^2 is negligible as compared with $[\ln(l/b_0)]^2$, the following approximate expressions for ζ_x, ζ_y can be obtained after substituting for ζ and ζ' from (40a) and (40b). We have

$$\left.\begin{aligned}&\zeta_x\cong\frac{(3-\Theta)}{\ln(l/b_0)+e_x};\quad \zeta_y\cong\frac{(3+\Theta)}{\ln(l/b_0)+e_y}\\ \text{with}\quad&e_x=\frac{2(1-\Theta)e_r+(1+\Theta)e_t}{3-\Theta}\\ &e_y=\frac{2(1+\Theta)e_r+(1-\Theta)e_t}{3+\Theta}\\ &\Theta=\frac{\sin2\chi_0}{2\chi_0}\end{aligned}\right\}, \tag{40c}$$

where e_r and e_t are given by (40a) and (40b) in terms of χ_0.

(d) In U_z motion, we have, from (34),

$$\left.\begin{aligned}&\zeta_z=\frac{4\pi\eta}{\ln(l/b_0)+e_z}\\ \text{where}\quad&e_z=\ln\frac{2\tan(\chi_0/4)}{\chi_0/4}+\frac{1}{2}\end{aligned}\right\}. \tag{40d}$$

As mentioned already, the variations of the shape of the particle from a straight ellipsoid to a closed ring can be defined by the angle χ_0. Now, as functions of χ_0, our results which are listed in (39) and (40) indicate that, in general, the frictional coefficient has a fractional form and furnishes two parameters: it has in the numerator a simple numerical factor m, representing the kinematic shape effect; in the denominator it has a parameter e of a more complicated character, representing the mutual hydrodynamical interaction between the different parts of the particle, either in the form of a cooperation of the parts in overcoming the resistance, or a disturbance by the bluntness of the shape. In general the dynamic shape factor e varies with shapes in all cases (compare Fig. 4), but the kinematic shape factor m varies only in the case of ζ_x, ζ_y, and not in the case of ζ_z (compare (40c), (40d), and Fig. 3). Therefore, in the case of a motion in the 0z direction, the variations of the frictional coefficient ζ_z reside only in e_z. It is interesting to notice that e_z

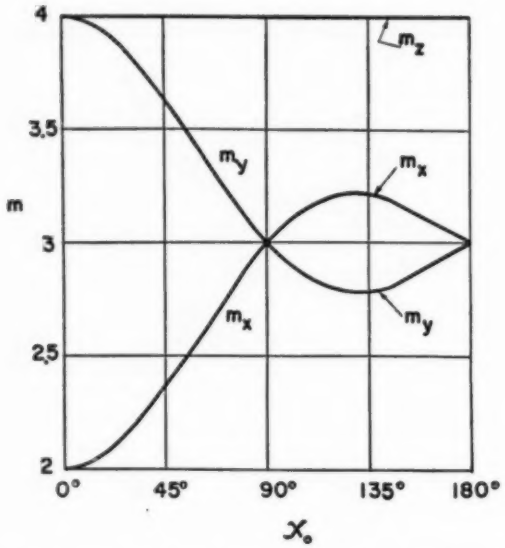


FIG. 3. Angular dependence of kinematic shape factor.

depends upon χ_0 in such a way that it increases as χ_0 increases (compare curve e_z of Fig. 4). This indicates a cooperative phenomenon of the liquid such that the particle in its translation perpendicular to its plane ($y0z$) has less resistance when it is closed than when it is opened. For motions in the other directions, such a cooperative phenomenon is not clearly manifested, because the nature of the frictional coefficients, as indicated by formulas (40c), is more complicated. For example, for the motion in the $0x$ direction, as the degree of bluntness increases by the increasing values of χ_0, the increase of the frictional coefficient is contributed partly by an increase in m_x, and partly by a decrease in e_x. The distinction between a half-circle and a ring may be of interest. They have a common value of m_x (see Fig. 3). Therefore, the larger bluntness in the ring will call for a smaller e_x in order to contribute a higher ζ_x. As illustrations, the expressions of the friction coefficients for the values of $\chi_0=0,\ \pi/2,\ \pi$, corresponding to a straight ellipsoid, a half-circle and a ring, are given in Table I. The expressions of resistances for a straight ellipsoid are identical with those found by Burgers,[7] and also with those values to be derived from the formula of Oberbeck.[6]

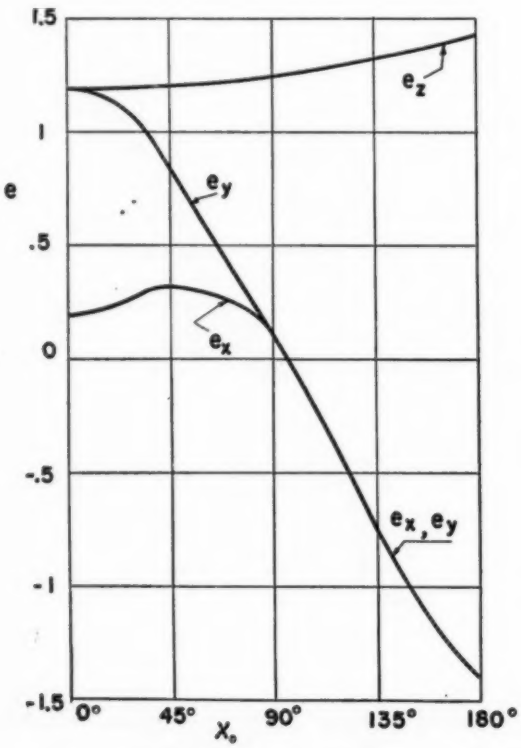


FIG. 4. Angular dependence of dynamic shape factor.

APPENDIX I

For the integrations of the equations of Oseen, we have need of a series of integrals which can be reduced to the four following types:

$$F_{2n}=\int_0^\nu d\nu\, S^{2n}\Delta^{-\frac{1}{2}},$$

$$G_{2n}=\int_0^\nu d\nu\, S^{2n}\Delta^{-\frac{1}{2}},$$

$$I_{2n}=\int_0^\nu d\nu\, S^{2n}SC\Delta^{-\frac{1}{2}},$$

$$J_{2n}=\int_0^\nu d\nu\, S^{2n}SC\Delta^{-\frac{1}{2}},$$

with Δ, S, and C defined by (5) and (7).

We have, first,

$$F_0=\int_0^\nu d\nu(\sin^2\nu+\epsilon^2)^{-\frac{1}{2}}.$$

Let us decompose this integral into two parts:

$$F_I=\int_0^{\epsilon^{\frac{1}{2}}} d\nu(\sin^2\nu+\epsilon^2)^{-\frac{1}{2}} \quad \text{and} \quad F_{II}=\int_{\epsilon^{\frac{1}{2}}}^\nu d\nu(\sin^2\nu+\epsilon^2)^{-\frac{1}{2}}.$$

To evaluate F_I put $\sin\nu=\nu-\frac{1}{6}\nu^3$ (ν^5, etc., are negligible, as being of the order $\epsilon^{5/2}$); then we form

$$\sin^2\nu+\epsilon^2\cong(\nu^2+\epsilon^2)-\tfrac{1}{3}(\nu^2+\epsilon^2)^2,$$

$$(\sin^2\nu+\epsilon^2)^{-\frac{1}{2}}\cong(\nu^2+\epsilon^2)^{-\frac{1}{2}}+\tfrac{1}{6}(\nu^2+\epsilon^2)^{\frac{1}{2}},$$

whence

$$F_I=\ln[\nu+(\nu^2+\epsilon^2)^{\frac{1}{2}}]+\tfrac{1}{12}\{\nu(\nu^2+\epsilon^2)^{\frac{1}{2}}+\epsilon^2\ln[\nu+(\nu^2+\epsilon^2)^{\frac{1}{2}}]\}\,|_0^{\epsilon^{\frac{1}{2}}}$$

$$\cong\ln\left\{\frac{2}{\epsilon^{\frac{1}{2}}}\left(1+\frac{\epsilon}{4}\right)\right\}+\frac{\epsilon}{12}\cong\ln\frac{2}{\epsilon^{\frac{1}{2}}}+\frac{\epsilon}{3}.$$

TABLE I. Frictional coefficients.[a]

	ζ	ζ'	ζ_x	ζ_y	ζ_z
Straight ellipsoid	$\frac{2\pi\eta}{\ln(l/b_0)+0.19}$	$\frac{4\pi\eta}{\ln(l/b_0)+1.19}$	$\frac{2\pi\eta}{\ln(l/b_0)+0.19}$	$\frac{4\pi\eta}{\ln(l/b_0)+1.19}$	$\frac{4\pi\eta}{\ln(l/b_0)+1.19}$
Half-circle	$\frac{2\pi\eta}{\ln(l/b_0)-0.30}$	$\frac{4\pi\eta}{\ln(l/b_0)+0.16}$	$\frac{3\pi\eta}{\ln(l/b_0)+0.01}$	$\frac{3\pi\eta}{\ln(l/b_0)+0.01}$	$\frac{4\pi\eta}{\ln(l/b_0)+1.25}$
Ring	$\frac{2\pi\eta}{\ln(l/b_0)-1.07}$	$\frac{4\pi\eta}{\ln(l/b_0)-1.57}$	$\frac{3\pi\eta}{\ln(l/b_0)-1.40}$	$\frac{3\pi\eta}{\ln(l/b_0)-1.40}$	$\frac{4\pi\eta}{\ln(l/b_0)+1.44}$

[a] ζ friction coefficient for a tangential motion; ζ' friction coefficient for a radial motion; ζ_x friction coefficient for a motion in the $0x$ direction; ζ_y friction coefficient for a motion in the $0y$ direction; ζ_z friction coefficient for a motion in the $0z$ direction; l half-length of the particle; $2b_0$ maximum thickness of the particle.

To evaluate F_{II} we shall use the expansion

$$F_{II}=\int_{\epsilon^{\frac{1}{2}}}^{\nu}\frac{d\nu}{(\sin^2\nu+\epsilon^2)^{\frac{1}{2}}}=\int_{\epsilon^{\frac{1}{2}}}^{\nu}\frac{d\nu}{\sin\nu}\left(1+\frac{\epsilon^2}{\sin^2\nu}\right)^{-\frac{1}{2}}$$

$$\cong\int_{\epsilon^{\frac{1}{2}}}^{\nu}\frac{d\nu}{\sin\nu}\left(1-\frac{\epsilon^2}{2\sin^2\nu}\right)=\ln\tan\frac{\nu}{2}+\frac{\epsilon^2}{4}\left(\frac{\cos\nu}{\sin^2\nu}-\ln\tan\frac{\nu}{2}\right)\Bigg|_{\epsilon^{\frac{1}{2}}}^{\nu}$$

$$\cong\ln\tan\frac{\nu}{2}-\ln\left\{\frac{\epsilon^{\frac{1}{2}}}{2}\left(1+\frac{\epsilon}{12}\right)\right\}-\frac{\epsilon}{4}\cong\ln\tan\frac{\nu}{2}-\ln\frac{\epsilon^{\frac{1}{2}}}{2}-\frac{\epsilon}{3}.$$

The addition of those two results yields

$$F_0=F_I+F_{II}=\ln\tan\frac{\nu}{2}+\ln\frac{4}{\epsilon}=\ln\left(\frac{4}{\epsilon}\tan\frac{\nu}{2}\right).$$

In order to calculate integrals $F_2, F_4, \cdots$, it is useful to remark that, in all integrals of the types

$$\int_0^{\nu}d\nu\sin^n\nu(\sin^2\nu+\epsilon^2)^{-m/2},\quad\int_0^{\nu}d\nu\sin^n\nu\cos\nu(\sin^2\nu+\epsilon^2)^{-m/2},$$

it is permissible to neglect ϵ^2 in the case of $n>m$. Hence, we obtain

$$F_2=1-\cos\nu;\quad F_4=\tfrac{2}{3}-\tfrac{3}{4}\cos\nu+\tfrac{1}{12}\cos3\nu;\quad\cdots.$$

The integrals G_{2n}, I_{2n}, J_{2n} can be calculated without difficulties either directly or with the same method as used for F_{2n}.

The calculations of the velocity perturbations require definite integrals taken between the limits

$$-(\chi_0+\chi)/2\quad\text{and}\quad(\chi_0-\chi)/2$$

for the variable ν, (i.e., between the limits $-\chi_0$ and $+\chi_0$ for χ'). Consider the definite integral

$$\int_{-(\chi_0+\chi)/2}^{(\chi_0-\chi)/2}d\nu\varphi(\nu).$$

It can be written as the sum of the two integrals as follows:

$$\int_{-(\chi_0+\chi)/2}^{(\chi_0-\chi)/2}d\nu\varphi(\nu)=\int_0^{(\chi_0-\chi)/2}d\nu\varphi(\nu)+\int_0^{(\chi_0+\chi)/2}d\nu\varphi(-\nu).$$

We notice that for F_{2n}, G_{2n} we can write $\varphi(-\nu)=\varphi(\nu)$, while for I_{2n}, J_{2n}, we can write $\varphi(-\nu)=-\varphi(\nu)$. Thus, it will be easy to calculate the definite integrals with the above limits by replacing ν in the indefinite integrals either by $(\chi_0-\chi)/2$ or by $(\chi_0+\chi)/2$, and to take the sum or the difference of the obtained expressions. For example, for the definite integral,

$$(F_0)=\int_{-(\chi_0+\chi)/2}^{(\chi_0-\chi)/2}d\nu\Delta^{-\frac{1}{2}},$$

we shall have

$$(F_0)=\ln\left\{\frac{16}{\epsilon^2}\tan\frac{\chi_0+\chi}{4}\tan\frac{\chi_0-\chi}{4}\right\},$$

which was denoted by σ in the text.

Growth of a Vapor Bubble in a Superheated Liquid*

H. K. FORSTER AND N. ZUBER
Department of Engineering, University of California, Los Angeles, California
(Received August 24, 1953)

The integro-differential equation for the growth of a vapor bubble in a superheated liquid is formulated and discussed. It is shown that two distinct time domains exist: one, of the order of 10^{-4} second, during which the effect of the hydrodynamic forces may be an important factor in the growth of the bubble, and another, during which this effect is unimportant. An integral equation is formulated for the latter domain. A solution of the problem, in closed form, valid for the entire interval of interest is presented; it agrees very well with experimental data for various superheats.

INTRODUCTION

IN problems of heat transfer with boiling, the time history of bubble formation and growth in a superheated liquid is of great importance. Lord Rayleigh[1] took the first step toward an understanding of the process of bubble growth (or collapse) when he formulated it as a problem in the hydrodynamics of an incompressible, inviscid fluid. Based on physical considerations of F. Bosnjakovic,[2] experiments were performed by Jakob[3] and co-workers. Their experiments, in which they measured the growth of a bubble by means of high-speed movies (500 to 1000 frames per second), yielded results in good agreement with Bosnjakovic theoretical predictions.

As liquid evaporates into the bubble, the surface of the bubble at which evaporation takes place moves. Various authors obtained solutions of the heat conduction equation for this problem; i.e., they solved the heat conduction equation for two phases with moving boundaries. The problem was first treated by Riemann and Weber[4] and recently by Carslaw and Jaeger,[5] Yoichi Mimura,[6] and others for a plane surface of separation between the two phases and by Plesset and Zwick,[7] to a first approximation for a spherical boundary. The latter mention in the closing paragraph of their paper that they have applied their approximate theory to the present problem, without, however, giving details or results.

FORMULATION AND DISCUSSION OF THE EQUATION

Consider Rayleigh's equation,[1]

$$\rho_2[R(d^2R/dt^2)+(\tfrac{3}{2})\dot{R}^2]=\Delta p, \qquad (1)$$

where ρ_2 is the density of the liquid, R the radius of the bubble, and Δp the difference of pressures inside and at great distance from the bubble. Rayleigh assumed constant Δp; but since heat is necessarily transferred to the bubble while it grows, the pressure Δp will obviously not remain constant but vary according to the laws of thermodynamics, i.e., Δp in Eq. (1) should be considered as variable. Moreover, it is well known from elementary physics that a fluid exerts a pressure $2\sigma/R$ (due to surface tension) inside a cavity of radius R and this term should therefore be subtracted from the pressure term Δp in Eq. (1). This gives the extended Rayleigh equation[8]

$$\rho_2[R(d^2R/dt^2)+(\tfrac{3}{2})\dot{R}^2]=\Delta p-(2\sigma/R). \qquad (2)$$

Equation (2) represents the laws of mechanics relating the force exerted by the bubble on the fluid (pressure) to the forces exerted by the fluid (surface tension and inertial forces). On the other hand, the pressure Δp in Eq. (2) is also determined by the thermodynamics of the process[9]: (1) The Clausius-Clapeyron equation connects the pressure difference to the temperature difference ΔT, i.e., to the difference in saturation temperatures inside and at great distance from the bubble (any spatial temperature variation within the bubble is neglected),

$$\Delta p=\frac{L}{T(v_1-v_2)}\Delta T, \qquad (3)$$

where L is the latent heat of evaporation, v_1 and v_2 the specific volumes of vapor and liquid, respectively,[10] and (2) the temperature difference is determined by the solution of the heat conduction problem with a moving surface of evaporation as mentioned in the Introduction. One may look upon Eq. (3) as a Taylor series for the pressure, broken off after the linear term.

* This study was supported in part by funds from the U. S. Atomic Energy Commission.

[1] Lord Rayleigh, Phil. Mag. **34**, 94 (1917).
[2] F. Bosnjakovic, Tech. Mech. Thermodynam. **1**, 358 (1930).
[3] M. Jakob, Z. Ver. dtsch. Ing. **76**, 1161 (1932).
[4] Reimann-Weber, *Die Partiellen Differential Gleichungen der Mathematischen Physik* (Vieveg and Sohn, Braunschweig, Germany) fifth edition, pp. 117–121.
[5] H. S. Carslaw and J. C. Jaeger, *Conduction of Heat in Solids* (Oxford University Press, London, England) first edition, p. 216 and following.
[6] Yoichi Mimura, J. Sci. Hiroshima Univ. **A14**, 127 (1950).
[7] M. S. Plesset and S. A. Zwick, J. Appl. Phys. **23**, 95 (1952).
[8] M. S. Plesset, J. Appl. Mech. **16**, 277 (1949).
[9] The suggestion that the variation of pressure in Rayleigh's equation should be obtained from thermodynamics was made by F. Romie, in a report, Liquid Systems Group, University of California, Los Angeles, California, 1951 (unpublished); see also Plesset and Zwick (reference 7).
[10] Numerical values for these quantities can be obtained (from tables) for a temperature T intermediate between T_∞ and T_0. For a possible improvement in this connection, see the remarks at the end of this paper.

For the present problem, the temperature difference can be written down directly starting from the principal solution of the heat-conduction equation in spherical coordinates. The liquid evaporating at the bubble wall constitutes a spherically distributed heat sink and the resulting temperature distribution is simply given by the well-known Green's function for the domain,[11] integrated over that space where evaporation took place,

$$\Delta T=-\frac{L\rho_1}{c_2(\pi a)^{\frac{1}{2}}\rho_2}\int_0^t\frac{R(t')\dot{R}(t')}{R(t)(t-t')^{\frac{1}{2}}}\times\left\{\exp\left[-\frac{(R-R')^2}{4a(t-t')}\right]-\exp\left[-\frac{(R+R')^2}{4a(t-t')}\right]\right\}dt'+(T_0-T_\infty)\quad(4)$$

where a is the thermal diffusivity, c_2 the specific heat; T_0 is the initial temperature of the system and T_∞ is the saturation temperature for the pressure p_∞ of the system (in other words, T_∞ and p_∞ are chosen as datum values). T_0-T_∞ in Eq. (4) is then the superheat which will subsequently be denoted by $\Delta\tau$ while the integral, for brevity, will be denoted by $J(t)$. The contribution of the convective term is unimportant in the present problem.

Equations (4) and (3) inserted into Eq. (2) yield the integro-differential equation governing bubble growth

$$R(d^2R/dt^2)+\tfrac{3}{2}\dot{R}^2+\frac{2\sigma}{\rho_2R}-\frac{L\Delta\tau}{\rho_2T(v_1-v_2)}+\frac{L^2\rho_1}{\rho_2^2T(v_1-v_2)(\pi a)^{\frac{1}{2}}c_2}J(t)=0.\quad(5)$$

The initial conditions are given by the initial radius and the initial radial velocity. If it is desired to consider the effect of heating after $t=0$, the appropriate function $F(t)$ has to be added to the right sides of Eqs. (4) and (5). The authors were particularly interested in the analysis of experiments carried out in the laboratories of this University with boiling liquids which are heated electrolytically by distributed heat sources.[12] In this particular case, the heat input Q per unit volume and time was constant and was applied to the liquid as a pulse for a short time; such heat input throws a term Qt/ρ_2c_2 to the right side of Eq. (4).

Let the coefficient of $1/R$ in Eq. (5) be denoted by α, the fourth term of the equation by β, and the coefficient of the integral by γ. Then the "critical radius" R_0 of the bubble, which is defined as that radius at which the bubble is in (unstable) equilibrium with its surroundings at a superheat $\Delta\tau$, is $R_0=\alpha/\beta$ and the initial conditions for the integration of Eq. (5) are $R=R_0$, $\dot{R}=0$, for $t=0$. It is advantageous to introduce the dimensionless function $r=R/R_0$; dividing the equation by R_0^2 and including the term Qt then gives

$$r\ddot{r}+(\tfrac{3}{2})\dot{r}^2-A(r-1)/r+BJ(t)=qt,\quad(6)$$

with

$$A=2\sigma/\rho_2R_0^3,\quad B=\gamma/R_0^2,$$

and

$$q=\frac{QL}{\rho_2^2c_2T(v_1-v_2)R_0^2}\qquad R_0=\frac{2\sigma}{\rho_0-\rho_\infty}.$$

It is useful to have in mind the orders of magnitude of the numerical values of the various quantities appearing in Eq. (6). More precisely, for water at atmospheric pressure and at a superheat of 3°C (i.e., $T_0=376$°A), the initial radius of the bubble $R_0=10^{-3}$ cm; r ranges from its initial value 1 to a few hundred (for a bubble which grows to a few millimeters); $A=1.2\times10^{11}$, $B=2\times10^8$ (both in cgs units).

The integral $J(t)$, if left in its general form as stated, represents an almost insurmountable obstacle to an analysis of the general behavior of solutions of Eq. (6). It is possible to show, however, that $J(t)$ may be greatly simplified, neglecting points of relatively minor importance while still retaining those features which are the determining ones in the present heat conduction problem.

The second exponential term under the integral is very small (compared to 1) in the entire interval of integration, because a is of the order 10^{-3} and the radius increases sufficiently fast (as will be seen in the sequel) so that the negative exponential remains small; the second exponential term may therefore be neglected. The first exponential term relates the heat evolved at the source $R(t')$ to the temperature at $R(t)$. An error is certainly committed when $R(t')$ is assumed equal to $R(t)$ in the exponential, but if the bubble is growing at a rapid rate, the error should be small. On the other hand, while the bubble is still growing slowly, that is at the start, $R(t')$ is in fact very nearly equal to $R(t)$ for the entire interval of integration. Consequently, the first exponential of $J(t)$ may be replaced by unity without serious error in the present problem.

The ratio $R(t')$ over $R(t)$ in the integrand is unity at the start and remains close to 1 as long as the bubble has not grown much in size. For later times the presence of the factor $(t-t')^{\frac{1}{2}}$ in the denominator weights preponderantly the contributions of the ingetrand for those time intervals where t' approaches t, that is when $R(t')$ approaches $R(t)$ and the foregoing ratio approaches 1. In the following preliminary discussion of the qualitative behavior of the solution of Eq. (6), $J(t)$ may therefor be replaced by the expression

$$I(t)=\int_0^t\frac{(dR/dt')dt'}{(t-t')^{\frac{1}{2}}}.\quad(7)$$

[11] See, for instance, reference 5, p. 219.

[12] Greenfield, Lipkis, Liu, and Zuber, "Studies on density transients in volume heated boiling liquids," Report 53-17 (CLN AECU 2529), July, 1953, Department of Engineering, University of California, Los Angeles, California.

It is understood that the factor $R(t')/R(t)$ will again be included in the integrand in Eq. (7) before proceeding to the calculation of the actual solution.

The substitution $t\cos^2 u = t - t'$ transforms the integral into one with time independent limits:

$$I(t) = 2t^{\frac{1}{2}} \int_0^{\pi/2} \dot{r}(t\sin^2 u)\sin u du, \tag{8}$$

and the inequality

$$\frac{r(t)-1}{t^{\frac{1}{2}}} < I(t) \leq 2t^{\frac{1}{2}}\dot{r}(t) \tag{9}$$

can easily be obtained. The lower bound is obtained by multiplying the integrand in Eq. (8) by $\cos u$ and subsequent integration. The upper bound is obtained by the consideration that $\ddot{r}$ is positive definite as long as $\dot{r}$ grows and the integral in Eq. (8) is therefore equal to $2t^{\frac{1}{2}}\dot{r}(\epsilon t)$ (where $0 \leq \epsilon \leq 1$) by the mean value theorem. Therefore, the lower bound holds for all time while the upper bound holds from $t=0$ for as long as $\dot{r}$ increases and for a certain time beyond that.

Consider the solution of Eq. (6) with heat input qt proportional to time in a neighborhood of $t=0$; $r(0)=1$ and $\dot{r}(0)=0$ and therefore $\ddot{r}(0)=0$. The integral may be developed into a series; one obtains from Eq. (8)

$$I(t) = \sum_{n=2}^{\infty} r_{(0)}{}^{(n+1)} \frac{n!\,2^{2n+1}}{(2n+1)!} t^{n+\frac{1}{2}}, \tag{10}$$

where

$$\int_0^{\pi/2} \sin u du = \frac{\sqrt{\pi}}{2} \frac{\Gamma[(n+1)/2]}{\Gamma[(n+2)/2]} \text{ was used.}$$

The leading term in the integral is of the order 5/2 in time; therefore the leading term in $\ddot{r}$ is qt and $r = 1 + qt^3/6$ in some neighborhood of $t=0$. It is then seen from Eq. (6) (in orders of magnitude), taking into account the numerical values for A and B, that $\dot{r}$ and $\ddot{r}$ are at first determined by qt, then, already after a few microseconds, by the superheat[13] terms; however, this term is seen to be balanced by the (negative) evaporation term within a fraction of a millisecond.

The problem may be studied further by considering Eq. (6) directly; neglecting qt which is unimportant after a short time and inserting numerical values for A and B which are important because they are so large,

$$r\ddot{r} + \tfrac{3}{2}\dot{r}^2 = 1.2\times 10^{11}\frac{r-1}{r} - 2\times 10^8 \int_0^t \frac{\dot{r}(t')}{(t-t')^{\frac{1}{2}}} dt'.$$

It can clearly be seen that the hydrodynamic problem which is here represented by the left side of the equation can be important in determining $\dot{r}$ and $\ddot{r}$ only while r is still of the order of 1. But, inserting either the upper or the lower bound for the evaporation term, it is seen that as soon as r has grown by an order of magnitude (which still takes place within a millisecond), the hydrodynamic terms have lost their importance; the tremendous values of the coefficients of the superheat and evaporation terms (10^8 and 10^{11}) cause $\dot{r}$ to be determined by the balance between these terms while the hydrodynamic terms with coefficients of the order of 1 may be neglected.

The complicated Eq. (6) need therefore be solved (if at all) only for a very short time; already after a millisecond r is much larger than 1 and certainly then, if not sooner, the much simpler integral equation

$$(r-1)/r = 1/CJ(t), \tag{11}$$

where $C = A/B$ should be considered. It must be stressed that, e.g., for $r=10$ the bubble has still only a radius of 10^{-2} cm so that Eq. (11) is in practice applicable during the entire "life" of the bubble.

The question arises whether at superheats as encountered in boiling liquids the hydrodynamic forces assume great importance at all, even in the beginning of bubble growth. This question may be answered by investigating whether neglect of the hydrodynamic forces for small bubble radii will bring about the appearance of great initial growth rates. More precisely, neglect of the hydrodynamic forces is equivalent to the assumption of a fluid with vanishing inertia. Does such an assumption have as a consequence the appearance of large or infinite initial rates of bubble growth? If this is not the case (and it will be shown in the sequel that it is not) and if the bubble grows slowly even when the hydrodynamic forces are neglected, it follows that the hydrodynamic forces are not important in this instance.

As long as $(r-1) \ll 1$, only the linear term in Eq. (11) need be considered which yields

$$\varphi(t) + q(t) = \frac{1}{C}\int_0^t \frac{\dot{\varphi}}{(t-t')^{\frac{1}{2}}} dt',$$

where $r = 1+\varphi(t)$ and $\varphi(0)=0$. This linear integral equation of the convolution type may be solved by taking the Laplace transform on both sides and has the solution $\varphi(s) = -Cq(s)/[C-(\pi s)^{\frac{1}{2}}]$, where $\varphi(s)$ and $q(s)$ are the respective transforms of $\varphi(t)$ and $q(t)$ in the usual notation. The solution for a distributed heat input $q(t) = q\cdot t$ is, therefore,

$$\varphi(t) = \frac{q}{\bar{c}^2}\left[\exp(\bar{c}^2 t)(1+\operatorname{erf}\bar{c}\sqrt{t}) - \bar{c}^2 t - \frac{2\bar{c}}{\sqrt{\pi}}\sqrt{t} - 1\right],$$

$$\varphi \ll 1, \quad \bar{c} = c/\sqrt{\pi}. \tag{12}$$

This solution is exact as long as φ^2 may be neglected compared to φ and describes the growth of the bubble in the beginning. It may be seen from Eq. (12) by inspection that even without consideration of hydrodynamic forces, the bubble starts its growth from critical size slowly, i.e., proportional to $t^{\frac{1}{2}}$.

[13] In the sequel, the third, fourth, and fifth terms in Eq. (6) will be called the superheat term, evaporation term and the heat-pulse term, respectively.

SOLUTION OF THE EQUATION AND CONCLUSION

Equation (6) was integrated numerically for a short time interval to verify the considerations presented above. Equation (10) was used in a step-by-step method which was laborious. For a short time interval in the neighborhood of $t=0$, the considerations presented above were checked by adapting the problem to the differential analyzer of the Department of Engineering of the University of California and the solutions obtained for superheats of 3°, 4°, and 5°C are presented in Fig. 1. The character of the solution is as was to be expected from the discussion. A sharp increase in r caused by the superheat term followed by a decrease of the increase brought about by the evaporation term which brings about an inflection point in $r(t)$ (i.e., $\dot{r}=0$).

The integral equation (11) which is here restated explicitly for convenience,

$$\frac{r-1}{r}+q(t)=\frac{1}{Cr}\int_0^t \frac{r(t')\dot{r}(t')}{(t-t')^{\frac{1}{2}}}dt',$$

is nonlinear, quasi-regular, of the Volterra type. One might obtain a solution numerically or by means of series, calculating the coefficients from the initial conditions in the usual way. However, as shown in the following lines, the solution may be obtained in a more direct manner.

First, an upper bound is found for the solution: by reasoning similar to that which yielded the inequality Eq. (9) it is found that the right side of the integral equation is larger than $(r^2-1)/2Crt^{\frac{1}{2}}$. If the inequality is then solved for r, it is seen that r remains smaller than $2Ct^{\frac{1}{2}}$ for all times except for an arbitrarily small neighborhood of $t=0$ because $q(t)$ may be chosen arbitrarily small. Now, since an explicit solution has already been obtained for small radii (Eq. (12)), one's attention turns naturally to the consideration of an asymptotic solution for large radii. But for large radii the (-1) and the arbitrarily small function $q(t)$ may be neglected compared to r and the solution of the resulting equation is seen by inspection to be simply $Ct^{\frac{1}{2}}$.

Neglect of (-1) compared to r is equivalent to neglecting surface tension and the error made in this manner is roughly equivalent to using a superheat which is in the ratio $1/r$ too high (i.e., 5 percent too high for $r=20$, 1 percent too high for $r=100$).

Consideration may now be given to the surface tension term $(1/r)$ and the initial growth rate which the bubble attained during the first few microseconds in the following manner: Let the integral in Eq. (11) be approximated by the function $2r\dot{r}t^{\frac{1}{2}}$ obtained from using the mean value theorem of the integral calculus. This value is at first too high but approaches the exact value of the integral as time increases. Let it therefore be used as an approximation for the integral in Eq. (11). This yields a simple differential equation of the first order which is solved by separation of variables,

$$r+\log\frac{r-1}{r_1-1}=Ct^{\frac{1}{2}} \quad t>\left(\frac{r_1}{C}\right)^2, \tag{13}$$

with initial condition $r=r_1$ for $t=r_1^2/C^2$. The initial radius r_1 may now be determined as that radius at which the rate of growth of the bubble resulting from the superheat term and the evaporation term attain equal order of magnitude. It was shown in the discussion of Eq. (6) that this radius is attained after a few microseconds while r is still of the order 1. Actual computation yields that the radius r has increased at that time from its original value of 1 by a few percent. (It may be mentioned that the precise value of this initial radius r_1 is of less influence on the solution than would appear at first because it enters logarithmically.)

A value $r_1=1.015$ was used in Eq. (13) and the results thus obtained are shown in Fig. 2 for various superheats. They are seen to be in good agreement with experimental values which are taken from a recently published paper by Dergarabedian.[14]

In studies on heat transfer and boiling, analytical expressions for bubble growth should prove very useful for the calculation of rates of evaporation, heat transfer coefficients, etc. From the foregoing it is suggested that Eq. (13) may be used as an implicit expression for the

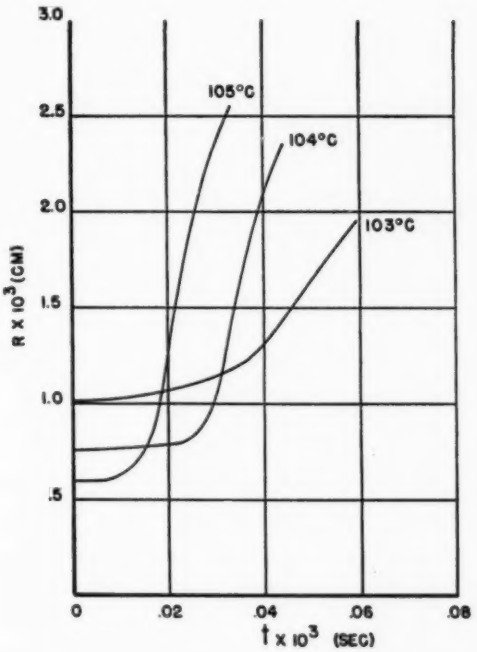


FIG. 1. Initial growth of a vapor bubble in a superheated liquid for various superheats.

[14] P. Dergarabedian, Paper No. 53-SA-10, Heat Transfer and Fluid Mechanics Institute, June, 1953, Los Angeles, California. (Dergarabedian includes in his graphs a curve obtained from Plesset and Zwick, otherwise unpublished, which agrees well with the results obtained here.)

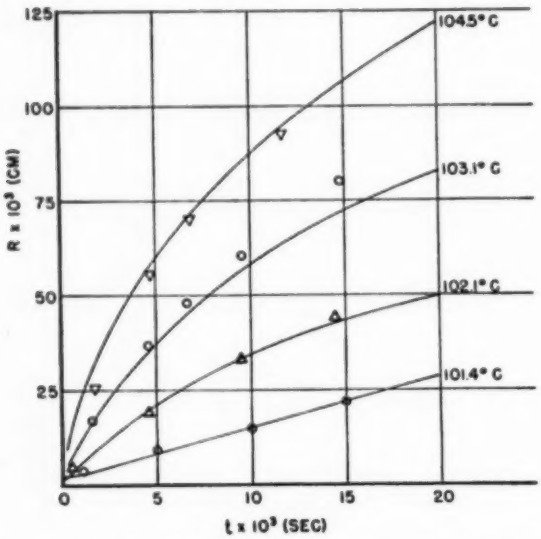


FIG. 2. Growth of a vapor bubble in a superheated liquid for various superheats from Eq. (13). [Circles and triangles show experimental values by Dergarabedian (reference 14).]

radius of a vapor bubble from the micron region upward, where

$$C=\frac{\Delta\tau(\pi a)^{\frac{1}{2}}c_1\rho_2}{R_0L\rho_1}. \tag{14}$$

It shall be stressed that the coefficient C which appears in Eqs. (13) and (14) is not a constant chosen *ad hoc* to fit certain data, but is computed from such thermodynamic coefficients of the fluid as the latent heat, density, etc. Thus, Eqs. (12)–(14) may be used, for instance, to obtain data on bubble growth at high pressures and temperatures for which experimental data are practically nonexistent.

In this connection, an improvement in the formulas is possible. Equation (13) may be used in Eq. (4) to obtain the temperature in the bubble as an explicit function of time. Then the more sensitive thermal constants which enter into the coefficient C (as for instance ρ_1) may be represented as (linear) functions of temperature. Since temperature was just obtained as a function of time from Eq. (4), the coefficient C will now be an explicit (slowly varying) function of time.

Concerning the limitations and possible extension of the theory as here presented: the spherically symmetric problem of heat conduction into a growing vapor bubble at rest has been solved in this article. As the bubble grows buoyant forces move it upward into surroundings whose temperature is higher because they have not been cooled by the withdrawal of heat of evaporation by the growing bubble. The final growth rate of the vapor bubble which moves upward in a fluid of constant temperature will therefore be higher than that of the stationary bubble here treated and will in time approach a constant. This constant growth rate is fairly well represented in current literature. If treated by a method similar to that used in this paper, the growth of a vapor bubble which moves because of buoyancy presents an axially symmetric problem of heat conduction with moving phase boundaries similar to, but more complicated than, the spherically symmetric problem here treated.

The authors wish to express their appreciation of the generous cooperation received from Professor L. M. K. Boelter, Dean of the College of Engineering, during this study.

Magnetic Switching Circuits

ROBERT C. MINNICK
Computation Laboratory of Harvard University, Cambridge, Massachusetts
(Received September 10, 1953)

Rectangular hysteresis-loop magnetic materials are widely used at present for the storage of digital information. The principles involved in static magnetic storage are generalized in this paper so that magnetic cores may be used for the purpose of switching. Magnetic circuits to produce the logical operations of "and" and "or" are described. The "and" and "or" circuits are combined into a class of circuits capable of producing any switching function of any number of input binary variables. All of the magnetic switching circuits described are devoid of rectifiers, vacuum tubes, transistors, and the like. Furthermore, these circuits are essentially independent of pulse amplitudes and durations.

MAGNETIC CORES AS STORAGE DEVICES

THE first storage device using magnetic cores was reported in 1948.[1] This device, called a *static magnetic delay line* or *magnetic shift register*,[2–4] consists of magnetic cores having hysteresis loops similar to that shown in Fig. 1. In the absence of an applied magnetic field, the core will be in one of states *a* or *b*, with which may be associated the binary values "0" and "1," respectively.

A magnetic storage device ordinarily has three windings as is shown in Fig. 2. A cycle of information storage and read-out in this core may be divided into three successive operations:

(1) *Reset.* Energize the control winding shown in Fig. 2. This applies a positive magnetic field to the core, after the completion of which the core is in state *a*.

(2) *Input.* The binary information to be stored, which has the form of a current pulse if this information is a binary "1" and no pulse if it is a "0," is applied to the input winding. After this operation the core in Fig. 2 is in state *a* if the input information is "0" and in state *b* if it is "1."

(3) *Advance.* To read this stored information out of the magnetic core, another current pulse is applied to the control winding. If the core is in state *a* ("0" is stored), this causes the hysteresis curve to be traced through cycle *a c a* (Fig. 1), and thus there is a very small change in flux. If, however, the core stands in state *b* ("1" is stored) when the control winding is energized, the hysteresis curve is traced through *b c a* and there is a correspondingly large change in flux. Thus the voltage induced across the output winding on Fig. 2 corresponds to the information which was stored in the core.

[1] Aiken, Wang, and Woo, "Static Magnetic Recording," Harvard Computation Laboratory *Progress Report No. 1, Investigations for the Design of Digital Calculating Machinery*, August 10, 1948.
[2] Kincaid, Alden, and Hanna, Electronics **24**, No. 1 (January, 1951).
[3] A. Wang, Various articles regarding "Static Magnetic Storage and Delay Line" contained in the Harvard Computation Laboratory *Progress Report Nos. 2, 3, 4, 5, 6, 8, 10, Investigations for the Design of Digital Calculating Machinery*, 1948, 1949, 1950.
[4] A. Wang, and W. D. Woo, J. Appl. Phys. **21**, 49 (1950).

ONE-LEVEL SWITCHING CIRCUITS

It is possible to generalize this magnetic core storage device into one which can perform binary switching. By a *switching device* is meant a circuit capable of accepting an arbitrary number of binary inputs (here in the form of the presence or absence of a pulse of current) and producing a specified binary function of these inputs. This output binary function is called a *switching function* of the input variables.[5] Consider the toroidal magnetic core shown schematically in Fig. 3(a). Each of the two "input windings" is energized by one of the input variables, x_1 and x_2. One winding is called the "reset winding," R; another is called the "advance winding," A. A fifth winding is for the output, $f(x_1,x_2)$. The arrows beside each winding denote the direction of magnetization of the magnetic core if there is current in the given winding.

Let the windings on the core in Fig. 3(a) be energized in the following order: reset, input, advance. Then upon application of the advance pulse, current will or will not

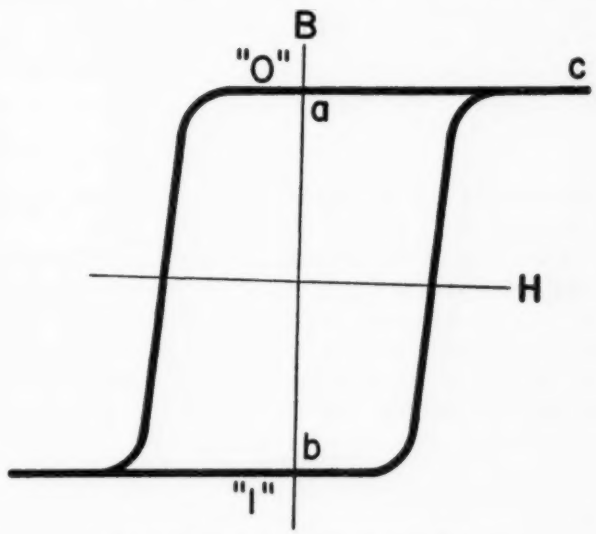


FIG. 1. Typical rectangular hysteresis loop.

[5] Staff of the Computation Laboratory, *Synthesis of Electronic Computing and Control Circuits*, Annals of the Computation Laboratory, Vol. XXVII, Harvard University Press, 1951.

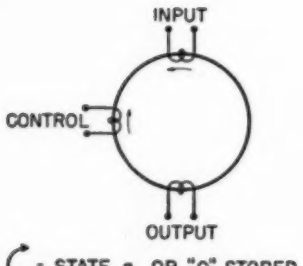


FIG. 2. Single magnetic core as a storage device.

be present in the output winding depending on the input currents. Suppose first that $x_1=x_2=0$. Then after the reset winding in Fig. 3(a) has been energized the core is magnetized in a clockwise direction. As both x_1 and x_2 are hypothesized to be "0," the state of the core is unchanged by the input operation. The advance pulse reverses the core from a clockwise to a counterclockwise state, and so there is a large change of flux which energizes the output winding. The output, $f(x_1,x_2)$, is assumed to occúr at the time that the advance pulse is applied, so in this case $f(x_1,x_2)=1$. Suppose now that x_1 or x_2 (or both) is "1." The reset pulse magnetizes the core in a clockwise sense, and as at least one of the input variables is "1," current will be present in one or both of the input windings; so during the input operation the core is magnetized counterclockwise. The advance pulse tends to magnetize the core in the same direction it is in, and so there is no appreciable change of flux in the core. Therefore in this case $f(x_1,x_2)=0$.

From this analysis, it is possible to list the binary output of this circuit for all four possible input conditions. This is shown in Table I. Defining the *inverse* of a binary variable, z, as $z'=1-z$, it is seen from this Table that the relation between $f(x_1,x_2)$ and the inputs is

$$f(x_1,x_2)=x_1'x_2'.$$

This circuit will be referred to as a one-level "and" circuit.

The circuit of Fig. 3(b) differs from that of Fig. 3(a) only in the reversed polarity of the advance winding. By a similar analysis to that given above, it is seen that the output of this circuit is $f(x_1,x_2)=1-x_1'x_2'$. An equivalent expression for this circuit may be obtained by the introduction of the *inclusive or* symbol, $\oplus$, which is defined by the relation

$$y\oplus z=1-y'z'.$$

Using this symbol, the output of the one-level circuit of Fig. 3(b) may be expressed as

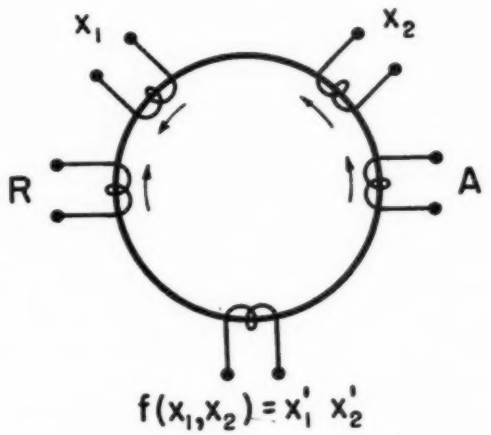
$$f(x_1,x_2)=x_1\oplus x_2.$$

This circuit will be referred to as a one-level "or" circuit.

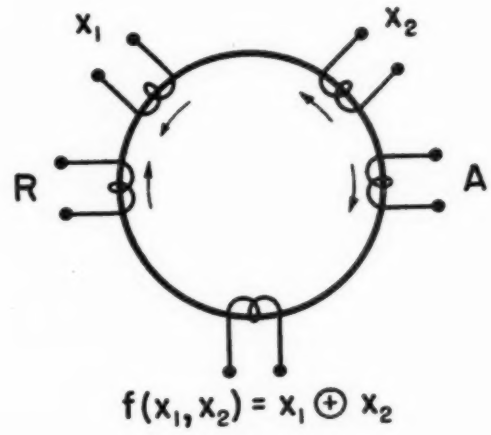
These two one-level magnetic core switching circuits may be generalized to the case of an arbitrary number of input variables. The circuits, along with the expressions for the output functions, are shown in Fig. 4. It should be noted that the circuit of Fig. 4(a) produces an output current pulse equivalent to the voltage output of n triodes connected with a single common plate resistor and having the voltages x_1, x_2, $\cdots$, x_n impressed on the grids. In a like manner, the magnetic core switching circuit of Fig. 4(b) is equivalent to n cathode followers connected with a single common cathode resistor.

TABLE I.

x_1	x_2	$f(x_1,x_2)$
0	0	1
0	1	0
1	0	0
1	1	0



(a)

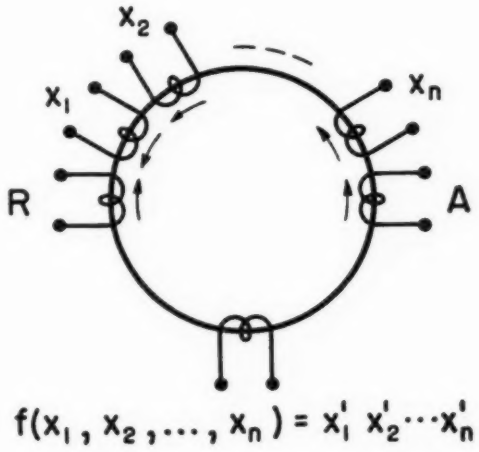


(b)

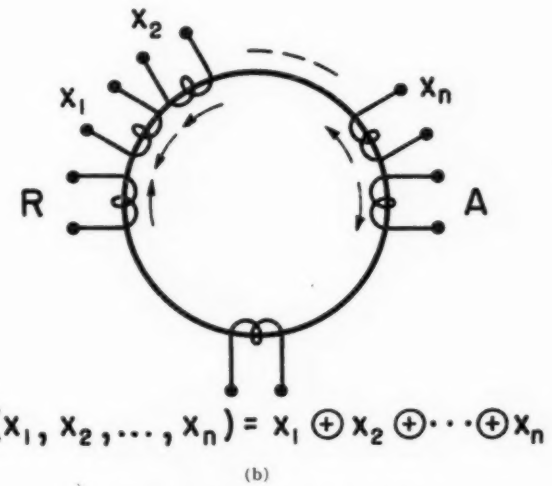
FIG. 3. One-level circuits of two variables: (a) "and" circuit; (b) "or" circuit.

TWO-LEVEL SWITCHING CIRCUITS

It is a widely known fact that by combining "and" and "or" circuits, any switching function of any number of input binary variables may be produced; therefore, it is desirable to find a means of interconnecting the one-level "and" and "or" circuits discussed above. It can be shown[6] that by the use of rectifier-resistance networks, identical with those used in static magnetic delay lines, it is possible to make such interconnections of the one-level "and" and "or" circuits. However, the advisability of using rectifiers in a magnetic switching circuit is questionable, for switching circuits can be constructed from rectifiers alone. In addition, the life of magnetic cores should be considerably greater than that of rectifiers; so a switching circuit composed only of magnetic cores, wire, and resistors should be extremely attractive from the point of view of lifetime.



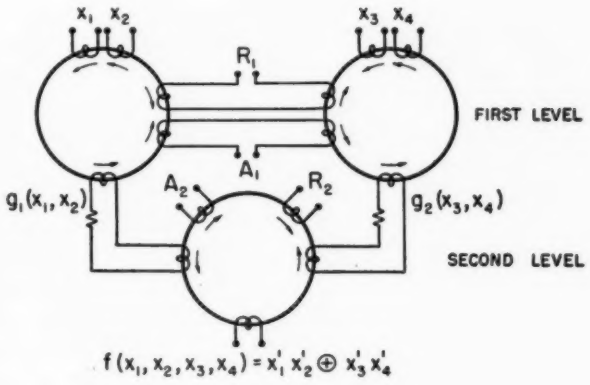
(a)



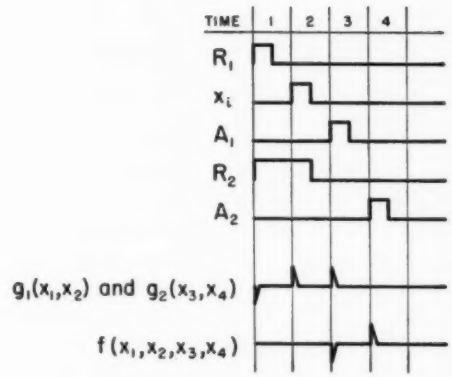
(b)

FIG. 4. One-level circuits of n variables: (a) "and" circuit; (b) "or" circuit.

[6] Minnick, Robert C., "The Use of Magnetic Cores as Switching Devices," doctoral thesis, Harvard University, April, 1953.



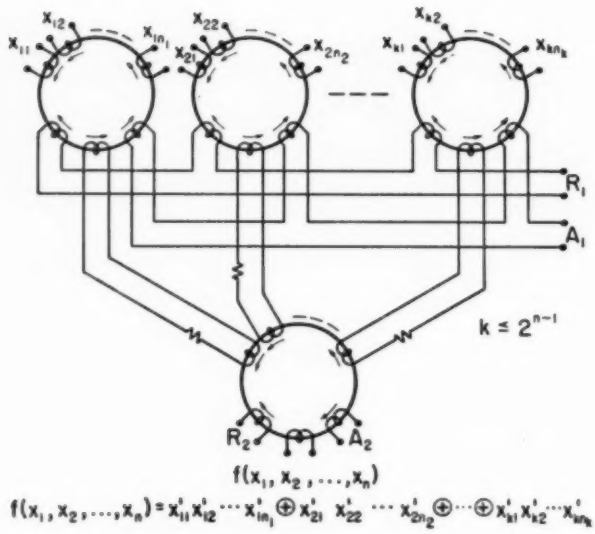
(a)



(b)

FIG. 5. Two-level circuit with two first-level cores: (a) schematic circuit; (b) timing pulses and waveforms.

A *two-level* magnetic core circuit containing no rectifiers is shown in Fig. 5(a). The four input variables, x_1, x_2, x_3, x_4 are introduced into the two *first-level* cores, and the outputs of these cores serve as inputs to the single *second-level* core. As is shown in the timing diagram in Fig. 5(b), the first-level cores are reset clockwise by the R_1 pulse during the first time interval. Following this, the inputs, x_1, x_2, x_3, x_4, are introduced. At the third time interval the A_1 pulse transfers the information in the first-level cores to the second level. The current wave forms on the output windings of the first-level cores are as shown in Fig. 5(b). The pulse which occurs at the third time interval is the information pulse, whereas the other two pulses are extraneous. The negative pulse which occurs during the first time interval appears to the second-level core as an additional reset. The pulse which occurs during the second time interval, however, is of the same polarity as the information pulse, so that some means of inhibiting the effect of this spurious pulse on the second-level core must be found. It is possible to remove this pulse by canceling it with a pulse of equal



$$f(x_1, x_2, \cdots, x_n) = x'_{11}x'_{12}\cdots x'_{1n_1} \oplus x'_{21}\, x'_{22} \cdots x'_{2n_2} \oplus \cdots \oplus x'_{k1}\, x'_{k2} \cdots x'_{kn_k}$$

FIG. 6. General two-level circuit.

magnitude but opposite sign; however, it is believed that methods of this sort should be avoided in that they involve the precise control of current amplitudes, an undesirable feature in a large scale computing system. This inhibition can be effected without canceling currents by the use of a double-length reset pulse, R_2, as is shown in Fig. 5(b). The current of the R_2 pulse is adjusted such that the magnetic field produced by the reset winding is greater than the largest possible field which can be produced by spurious pulses. In this way the second-level core remains in its reset state at the end of the second time interval, and the only pulses which affect the second-level core are the first-level outputs, $g_1(x_1,x_2)$ and $g_2(x_3,x_4)$, which occur during the third time interval. By comparing Figs. 5(a) and 3(a), it is clear that the first-level outputs are

$$g_1(x_1,x_2) = x_1'x_2'$$
$$g_2(x_3,x_4) = x_3'x_4'.$$

The second-level core operates similarly to the one in Fig. 3(b), the only difference being the double-length reset pulse, R_2, in the two-level circuit. Therefore, the final output function in the circuit of Fig. 5(a) is

$$f(x_1,x_2,x_3,x_4) = g_1(x_1,x_2) \oplus g_2(x_3,x_4) = x_1'x_2' \oplus x_3'x_4'. \quad (1)$$

The wave form of this final output function is shown in Fig. 5(b). It is seen that the information pulse is the only positive pulse which appears at the output. Thus the output winding of such a two-level circuit may be connected directly to the input of another similar circuit with no intermediate gating devices to separate the information pulse from extraneous pulses. In addition, it has been shown[6] both by experiment and by theoretical studies that it is possible to transfer the information from the output winding of a two-level circuit simultaneously to the input windings of at least two other similar two-level circuits with no decrease in the signal amplitude. It seems possible, therefore, to construct from magnetic cores a complete switching system which contains no intermediate amplifiers.

The most general two-level circuit is shown in Fig. 6. Each of the input variables, x_{ij}, in this circuit are defined as one of x_1 or x_1', x_2 or x_2', $\cdots$, x_n or x_n'. The maximum number of first-level cores necessary to produce any switching function of n input variables can be shown never to exceed 2^{n-1}.

EXPERIMENTAL RESULTS

Typical magnetic core switching circuits were constructed in order to verify the properties described above. The core material used in these experiments was two and one-half wraps of 0.001-in.×0.125-in. Deltamax tape enclosed in a case of 0.75-in. outside diameter. These are the same cores as are used in the fast storage delay lines in the Harvard Automatic Magnetic Drum Calculator (Mark IV).

In the first experiment a single magnetic core was

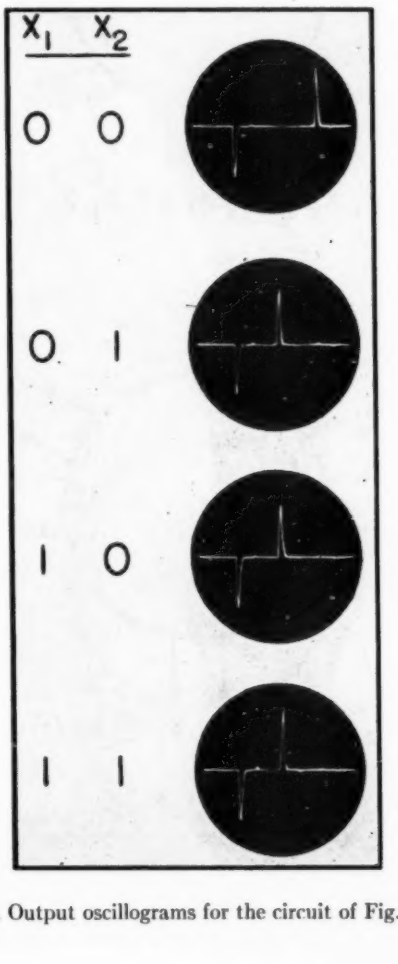


FIG. 7. Output oscillograms for the circuit of Fig. 3(a).

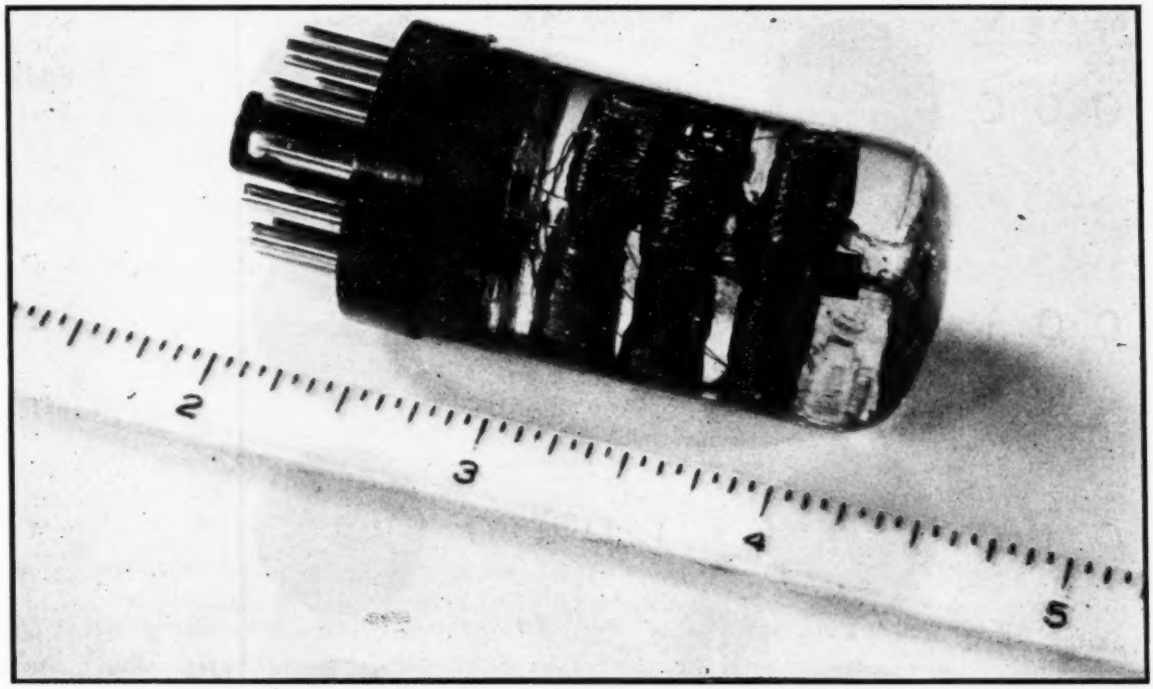


FIG. 8. Cast two-level circuit with two first-level cores.

connected in the one-level circuit of Fig. 3(a), and oscillograms of the current waveforms were recorded for the four possible input combinations. On these oscillograms, shown in Fig. 7, the first pulse is the effect of the reset operation; the second pulse is due to the inputs; and the third pulse is the actual information. The duration of the entire operation shown in Fig. 7 is approximately 70 microseconds.

The two-level circuit of Fig. 5(a) also was constructed from Deltamax cores. This circuit, cast in plastic and mounted on a standard tube base, is pictured in Fig. 8. Oscillograms are shown in Fig. 9 for all sixteen combinations of the input variables. It is seen from this figure that the only positive pulse represents the output information, whereas the negative pulse is due to the first-level advance, A_1. The over-all operation time for this two-level circuit is approximately 100 microseconds.

CONCLUSIONS

It was found in the above experiments that the results were nearly unchanged by variations which normally would be expected in the magnetic material. The behavior of these switching circuits was found to be essentially independent of the current amplitude and of the pulse duration, above certain minimum values. The pulse shape was found to be of little importance, except that the rise time had to be kept greater than a minimum value in order to prevent excessive eddy current losses.

There is no extremely sensitive parameter in these magnetic switching circuits which must be adjusted carefully in order to make them operate properly. This property of the circuits is of considerable importance if they are to be used in large scale computing systems; for in such systems many individual switching circuits are used, and in order to maintain a reasonable over-all reliability it is necessary to have wide tolerances in the individual switching circuits.

The experiments described above were performed with the object of demonstrating the logical properties of magnetic switching circuits. As the logical properties of a switching device are, within reason, independent of speed of operation of the circuit, no special effort was made to reduce the operation time to a minimum. The high eddy current losses in the metal ribbon material used in these experiments requires that operation times be kept fairly long. To achieve faster switching it is clear that fractional-mil metal ribbon or magnetic ferrite material is indicated. It seems reasonable to expect that reliable magnetic core switching circuits can be built, from materials available at the present, to operate in one to five microseconds.

ACKNOWLEDGMENTS

The author wishes to express his sincere appreciation to Professor Howard H. Aiken, who proposed this research on magnetic switching circuits and who has offered many valuable suggestions throughout the course of the work. The advice of several other members

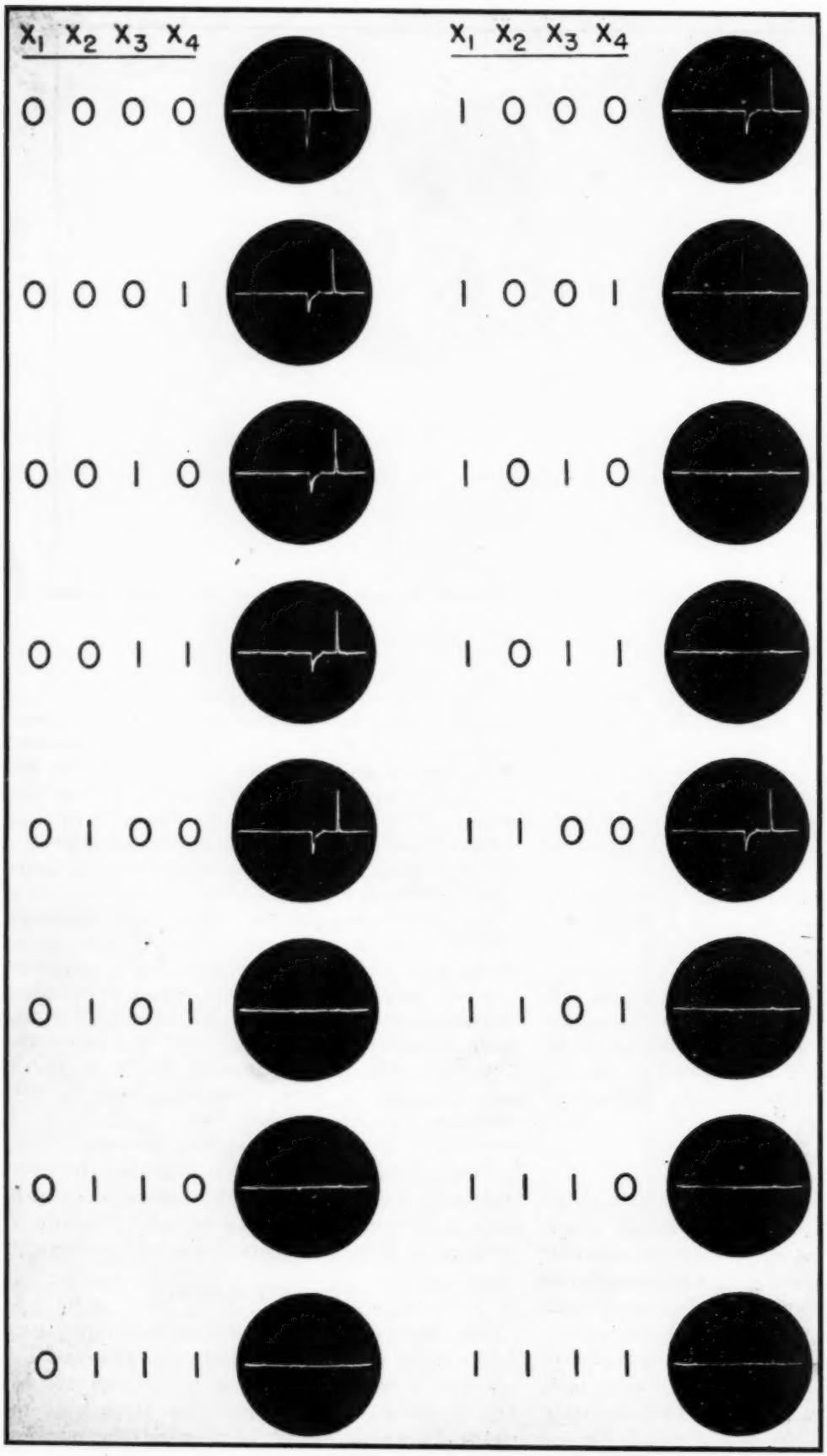


FIG. 9. Output oscillograms of the circuit of Fig. 5.

of the Staff of the Computation Laboratory of Harvard University, especially that of Mr. Kenneth E. Perry and Mr. Kenneth E. Iverson, is gratefully acknowledged.

This research was supported in part by the Bell Telephone Laboratories through a contractual agreement with Harvard University. The author is indebted to this organization for its cooperation and support.

JOURNAL OF APPLIED PHYSICS VOLUME 25, NUMBER 4 APRIL, 1954

A Graphical Method of Evaluating Trigonometric Functions Used in Crystal Structure Analysis. Parts I and II*

H. J. GRENVILLE-WELLS†
Laboratory for Insulation Research, Massachusetts Institute of Technology, Cambridge, Massachusetts
(Received August 31, 1953)

A graphical method is described for evaluating structure factors of the form

$$F(hk0)=\sum_{0}^{n} f_r(hk0)\begin{matrix}\sin\\\cos\end{matrix}2\pi(hx_r+ky_r)$$

and electron density expressions of the form $\rho(X,Y)=\Sigma\,\Sigma\,F(hk0)\begin{matrix}\sin\\\cos\end{matrix}2\pi(hX+kY)$. The method depends upon the fact that if, in a [001] projection of a unit cell the normal to the planes $(hk0)$ is graduated with the appropriate sine or cosine function giving the phases of points along the normal, then a circle with its center at $(\frac{1}{2}x,\frac{1}{2}y)$, passing through the origin, will cut the normal to $F(hk0)$ at the value of $\begin{matrix}\sin\\\cos\end{matrix}2\pi(hx+ky)$. Part I contains the basic principles and a standard chart, and Part II cites examples of the use of the method and the accuracy to be expected in structure factor calculations, and of its application to least-squares refinements.

I. BASIC PRINCIPLES AND STANDARD CHARTS

1. Introduction

THE construction discussed below provides a rapid graphical method for evaluating trigonometric functions of the form $\begin{matrix}\sin\\\cos\end{matrix}2\pi(hx+ky)$, or functions such as $\begin{matrix}\sin\\\cos\end{matrix}2\pi hx\cdot\begin{matrix}\sin\\\cos\end{matrix}2\pi ky$, which can be reduced to this form.

It is primarily intended to apply to two-dimensional projections of a crystal structure, and the nomenclature used refers to a [001] projection. It can be used either on the true projection of the unit cell, or with the atomic coordinates referred to a square projection, a procedure which has advantages to be discussed in Sec. 4. The real and not the reciprocal lattice is used. The arguments developed below for cosine functions should be understood to apply to sine functions also.

2. Basic Construction for Evaluating Structure Factors

Consider the [001] projection of the unit cell (Fig. 1). AB is the trace of the plane $(hk0)$ so that $OA=a/h$, $OB=b/k$. OP is the trace of the plane through O parallel to AB. The contribution to the structure factor $F(hk0)$ from the atom at (x,y) can be obtained from a Bragg-Lipson chart,[1] and the line PQ can be graduated from such a chart. The particular utility of this construction lies in the fact that since PQ must be perpendicular to OP, points such as P will lie on the circle OPQ for all planes $(hk0)$. This is the circle with its center at $(\frac{1}{2}x,\frac{1}{2}y)$ which passes through the origin. ON is the normal through the origin to the planes $(hk0)$, and it can be seen that OR is equal and parallel to PQ. Hence, if ON instead of PQ were graduated from the appropriate Bragg-Lipson chart, the value of $\cos2\pi(hx_r+ky_r)$

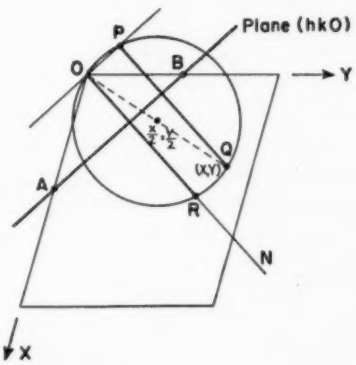


FIG. 1. Basic construction for structure factors of the form

$$F(hk0)=\sum_{0}^{n} f_r(hk0)\cos2\pi(hx_r+ky_r).$$

---

* Sponsored by the U. S. Office of Naval Research, the U. S. Army Signal Corps, and the U. S. Air Force under ONR Contracts N5ori-07801 and N5ori-07858.

† Present address: Chemistry Department, University College, London W.C.1., England.

[1] W. L. Bragg and H. Lipson, Z. Krist. **A95**, 323 (1936).

FIG. 2. Example of construction for evaluating structure factors.

would be the reading at the point where the circle with center $(\frac{1}{2}x_r,\frac{1}{2}y_r)$ intersects ON. ON will be called the "$F(hk)$ line."

Generalizing the construction, it can be seen that for any $(hk0)$ plane, the geometrical structure factor $F(hk0)$ will be the sum of the readings on the corresponding $F(hk)$ line where it is intersected by the n circles corresponding to the n atoms in the projection; i.e.,

$$F(hk0)=\sum_0^n \cos 2\pi(hx_r+ky_r).$$

The construction is illustrated in Fig. 2. If the sine or cosine function is multiplied by the appropriate value of the atomic scattering factor $f_r(hk0)$, then the crystal structure factor

$$F(hk0)=\sum_0^n f_r(hk0)\cos 2\pi(hx_r+ky_r)$$

will be obtained.

3. Basic Construction for Evaluating Electron Density

By analogy with the preceding construction, if a set of normals to the planes $(hk0)$ are graduated as before, but instead of multiplying them by the value of the atomic scattering factor, they are multiplied by the *observed values of* $F(hk0)$ (see Fig. 3), a circle whose center is at $(\frac{1}{2}X,\frac{1}{2}Y)$ will cut the $F(h_1k_1)$ line at A, which gives the value of $F(h_1k_1)\cos 2\pi(h_1X+k_1Y)$, and so on. Hence the sum of the readings on the circle

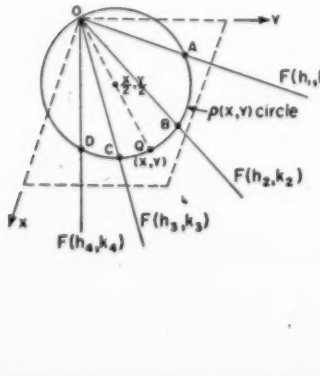


FIG. 3. Basic construction for electron density summations of the form

$$\rho(X,Y)=\sum_0^H\sum_0^K F(hk0)\times\cos 2\pi(hX+kY).$$

$OABCD$, which will be called the "$\rho(X,Y)$ circle," will be $F(h_1k_1)\cos 2\pi(h_1X+k_1Y)+F(h_2k_2)\cos 2\pi(h_2X+k_2Y)+\cdots$ for all the observed values of $F(hk0)$; i.e., the sum will be $\sum_H\sum_K F(hk0)\cos 2\pi(hX+kY)$ which is the value of the electron density $\rho(X,Y)$ at the point (X,Y).

4. Use of a Square Projection

There are three main advantages in the use of a square projection: (1) The directions and the graduations of the $F(hk)$ lines will be the same for all crystals, so that standard projections can be prepared instead of a new set being necessary for each projection which is not already square, and only the numerical values of $f(hk)$ or $F(hk0)$ need be inserted as required. (2) Since the [110] direction is a mirror line, only the octant for which, say, $k\geqslant h$ need be drawn out. (3) The $F(hk)$ lines can be graduated with very little labor by making use of the construction shown in Fig. 4; if a circle drawn with its center at the point $(\frac{1}{2},\frac{1}{2})$ intersects the $F(hk)$ line at P, OP has to be divided into $(h+k)$ parts; i.e., in Fig. 4, $F(4,2)$ is divided into $4+2=6$ parts. If the

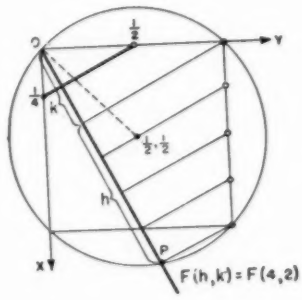


FIG. 4. Graduation of $F(hk)$ lines using a square projection.

charts are drawn on tracing paper, a piece of graph paper can conveniently be used as a proportional divider to graduate the $F(hk)$ lines.

In this case the crystal projection is distorted, but this does not appear to be a serious disadvantage.

5. Extension to Structure Factors of the Form

$$F(hk0)=\sum_0^n \begin{matrix}\sin\\\cos\end{matrix} 2\pi hx_r\cdot\begin{matrix}\sin\\\cos\end{matrix} 2\pi ky_r$$

The expression $2\begin{matrix}\sin\\\cos\end{matrix}2\pi hx\cdot\begin{matrix}\sin\\\cos\end{matrix}2\pi ky$ can be rewritten in the form $\cos 2\pi(hx-ky)\mp\cos 2\pi(hx+ky)$. This is equivalent to placing one atom at the point (x,y) and another at $(x,\bar{y})$. Then in Fig. 5, the value of $2\cos 2\pi hx\cdot\cos 2\pi ky$ will be the sum of the readings at A and B. Although an atom at $(x,\bar{y})$ is actually located at A (Fig. 6) so that its contribution to $F(hk)$ occurs at P, it can be seen that if $O'Q$ is drawn parallel to OP, in which case $O'Q=OP$, the reading at Q will be equivalent to the reading at R, since BR is necessarily perpendicular to OR, and hence to $O'Q$, and therefore the atom at $(x,\bar{y})$ can be placed at B and the construction confined

to one quadrant, as in Fig. 5. (It should be noted that for atoms having x or y zero, only one circle will appear.) This is a convenient fact, because the equivalence of identical positions in neighboring unit cells means that if either or both of the coordinates x and y are so small that the resulting circle is nearly tangential to some $F(hk)$ lines, with correspondingly increased errors in the positions of the intersections, the atom can be transferred to the next cell so that more nearly orthogonal intersections are obtained. The standard charts extend further than the single-cell range to permit a limited amount of such relocation. It should be noted that a circle will be truly tangential to an $F(hk)$ line only if (x,y) lies on the line through the origin perpendicular to the $F(hk)$ line, i.e., if (x,y) lies in the $(hk0)$ plane. Hence it will make a full contribution to $F(hk0)$.

6. Cases Where $hk\bar{0}$ Reflections Differ Numerically from $hk0$ Reflections

The construction as so far discussed has been confined to the case where h and k are both positive. For non-rectangular projections and in other cases where [010] or [100] is not a mirror line, $F(h\bar{k})$ values will not be

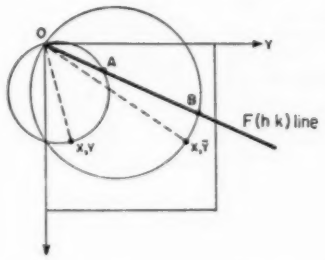


FIG. 5. Construction for structure factors of the form

$$F(hk0)=\sum_{0}^{n} f_r(hk)\cos 2\pi hx_r \cdot \cos 2\pi ky_r.$$

numerically equivalent to $F(hk)$ values, and will have to be obtained separately. But since

$$F(h\bar{k})=\sum_{0}^{n}\cos 2\pi(hx_r-ky_r),$$

these values can be obtained by locating atoms at $(x,\bar{y})$ instead of at (x,y), and reading the intersections on the $F(hk)$ lines.

7. Standard Charts

Standard charts have been prepared for $\frac{\sin}{\cos}2\pi(hx+ky)$ on the basis of a unit cell projection 10 in. square. The charts now completed cover only the 48 reflections for which h, $k\leqslant 6$ because this particular group is now being studied in connection with a least-squares technique. There are 3 cosine charts and 3 sine charts, each covering 16 reflections. It would be quite possible to place about 30 reflections on one such 10-in. chart, but the practical difficulty of showing more than two orders of $F(hk0)$ on the same line means that unless some specially modified chart were made to accomodate $F(h00)$ and $F(hh0)$ separately, a new chart is needed for each

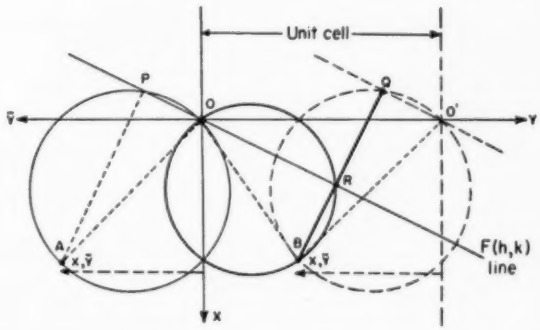


FIG. 6. Equivalence of atomic positions $(x,\bar{y})$ and $(x,1-y)$.

pair of values $h=1, 2$; $h=3, 4$, etc. The 48 reflections have therefore been distributed over the three charts so as to be as evenly spaced as possible. A specimen chart is reproduced in Fig. 7. On the original charts the negative sections of each $F(hk0)$ line have been marked in red for convenience in reading. The charts are graduated at 15° intervals, but the numerical values have been omitted except in the key to each chart, since it was found that after a few minutes' practice with the charts these numbers had been memorized, and the values at intersections readily interpolated. Furthermore, if such charts are reproduced, it is then a simple matter to take a fresh one for each new projection and insert the appropriate atomic scattering factor values or observed $F(hk0)$ values.

If the atomic coordinates are plotted on transparent paper in a cell 5 in. square (done in practice by superposing the tracing on a sheet of half-inch graph paper) these points will be the centers of the required circles, since the (x,y) circle has its center at $(\frac{1}{2}x,\frac{1}{2}y)$.

In large unit cells, the first 48 reflections represent a correspondingly reduced percentage of the observable $F(hk0)$ reflections, but in the case of projections of high symmetry in which coordinates need be sought only, say, in the range $x,y=0$ to 0.25 or 0.50, values for higher orders of any of the planes $F(hk0)$ already present on the charts can be obtained by drawing the (x,y) circles on a larger scale. For it was shown in Fig. 4 that the length OP of the $F(hk)$ line had to be divided into $(h+k)$ parts, so that the $F(4,2)$ line for a unit cell of side 10 in. will be the $F(8,4)$ line for a unit cell of side 20 in.

A *complete* set of $F(hk0)$ values cannot, of course, result, since any plane such as $F(9,8)$ which is a first-order reflection for which h or $k>6$ cannot be obtained in this way. This is a disadvantage when electron density maps are to be prepared from a Fourier series, but not necessarily so in the case of a least-squares refinement, for which a complete set of $F(hk0)$ values is not necessary.

8. General Utility of the Construction

The construction appears to have some value in the fact that it offers a very easy way of visualizing the

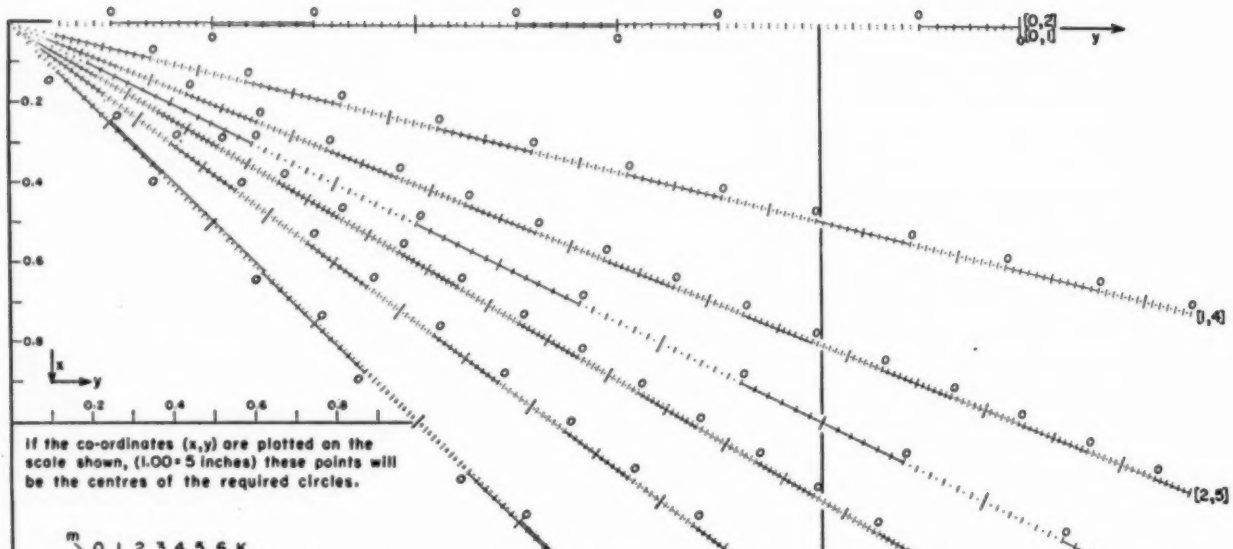

[0,4]
[2,5]
[0,2]
[3,5]
[3,4]
[0,0]
[2,2]
If the co-ordinates (x,y) are plotted on the scale shown, (1.00=5 inches) these points will be the centres of the required circles.
CHART FOR EVALUATING COS (hx + Ky) FOR THE VALUES OF h AND K INDICATED HERE.
x = values obtained directly
✓= values obtained by interchanging x and y: i.e. by rotating the paper carrying the (x,y) circles about the line mm
Scales are graduated at 15° intervals:
1.00 .96 .86 .70 .50 .25 0

FIG. 7. One of the standard charts. The originals are based on a 10-in. square.

contribution from any atom or group of atoms to a large number of structure factors simultaneously. Space-group extinctions can also be readily represented, as, for example, in Fig. 8. Consider the face-centered [001] projection in which the atom at the origin is represented by the circle through A, and the atom at the face center by the circle through B. These circles cut the $F(hk)$ line at P and Q, respectively, and $OQ=\frac{1}{2}OP$. In Sec. 3 it was shown that OP has to be graduated so as to have $(h+k)$ equal parts with maxima at O and P. Hence, if $h+k=2n$, Q will lie on a maximum, while if $h+k=2n+1$, Q will lie on a minimum, and atom B will cancel out atom A's contribution. Hence the space-group condition for a centered (001) projection can be derived; namely, reflections are present only for $h+k=2n$.

II. APPLICATION TO STRUCTURE FACTOR CALCULATIONS AND TO THE LEAST-SQUARES REFINEMENT PROCESS

Some practical applications of the graphical method described in Part I of this paper are given below. The standard charts described in Part I, Sec. 7 have been used in the calculations.

1. Structure Factors of the Form $\cos 2\pi(hx+ky)$

This is the form of equation for which the method is best suited; the contribution of each atom to a structure factor is the value read at a single intersection, as described in I, Sec. 2. As an indication of the accuracy obtainable with charts drawn on the 10-in. scale, values for 30 geometrical structure factors $[F(h0l)_{\text{graph}}]$ for the 12 carbon atoms in the two phenyl groups in a trial structure of diphenyl sulfoxide are given in Table I, together with the values $[F(h0l)_{\text{calc}}]$ computed for h, $l\leqslant 10$ on I.B.M. punched-card machines from coordinates given to the nearest 0.001 of the cell edge. The formula used was

$$F(h0l)=\sum_{r=1}^{12}\cos 2\pi(hx_r+ky_r)$$

and hence $F_{\max}[=F(000)]=12$. The space-group of diphenyl sulfoxide is $P2_1/n$, so that in the $h0l$ projection, reflections are present only for $h+l=2n$.

The 30 reflections are of three types: (1) $h0l$ reflections for which $0\leqslant h,\ l\leqslant 6$ (see I, Sec. 2). (2) This is a case to which the procedure described in I, Sec. 6 is

applicable, since $F(h0l)$ is not numerically equal to $F(\bar{h}0l)$. (3) It is also a case in which the projection is centered, so that for every atom at (x,y) there is an atom at $(\frac{1}{2}+x, \frac{1}{2}+y)$, which means in effect that either the x or the y coordinate need be sought only in the range 0 to 0.5; in this particular trial structure all x and y coordinates are small enough so that those reflections given in the standard charts can be used to obtain second orders, as described in I, Sec. 7, by plotting the (x,y) positions in a 20-in. unit cell instead of a 10-in. cell.

2. Structure Factors of the Form $\begin{matrix}\sin\\ \cos\end{matrix} 2\pi hx \cdot \begin{matrix}\sin\\ \cos\end{matrix} 2\pi ky$

This case is slightly less favorable than the preceding one, because the contribution from each atom is obtained from the readings at two intersections instead of at only one. In the case of cos·cos expressions, in which the *sum* of two readings is required, the procedure is exactly as in Sec. 1 above, except that there are effectively twice as many atoms; but for sin·sin functions, when the *differences* are needed, care should be taken to keep them separated (for instance, by drawing the (x,y) circles on one tracing and the $(x,\bar{y})$ circles on another).

As a further illustration of the accuracy to be expected, a six-parameter example which was being studied in connection with a least-squares technique, and which had been worked out on a hand-calculating machine to the nearest 0.001, was reworked using the standard charts. Twenty-four principal structure factors (for h, $k=0$ to 4) were calculated both for the postulated structure ("F_{obs}") and a trial structure ("F_{calc}"). The discrepancy value $([\sum|F_{calc}|-|F_{obs}|]/\sum|F_{obs}|)$ for the graphical method was $R_{graph}=33.7$ percent compared with $R_{calc}=35.8$ percent. The plane group chosen was *pmm*, so that in this case the form of the structure factor was

$$F(hk0)=\sum_{r=1}^{3} 4\cos 2\pi hx_r\cdot\cos 2\pi ky_r,$$

but a factor of 2 has been dropped throughout, so that in effect $F_{max}[=F(000)]=6$.

Comparing the values obtained with those given in Table I, for which $F_{max}=12$, as compared with $F_{max}=6$ in this example, the errors were found to represent a similar percentage of F_{max} in both cases. It has been found in practice that readings with the same set of circles are, in general, reproducible with much smaller differences than the worst discrepancies shown between F_{calc} and F_{graph} in Table I. This suggests that the errors are mainly due either to inaccuracies in the graduations of the standard charts, or more particularly to errors in the positions of the circles, rather than to errors made when interpolating the readings between the graduations; hence geometrical structure factors accurate to about 2 to 3 percent of F_{max} should be obtainable even with charts drawn on the relatively small 10-in. scale, if reasonable care is taken in constructing them.

TABLE I. Comparison between graphical and calculated values for some geometrical structure factors associated with a group of 12 similar atoms.

	Type (1)		Type (2)			Type (3)	
$h0l$	F_{graph}	F_{calc}	F_{graph}	F_{calc}	$h0l$	F_{graph}	F_{calc}
101	+0.68	+0.87	+2.19	+2.20	208	+1.56	+1.53
301	−3.13	−3.02	−0.34	−0.53	2, 0, 10	+1.35	+1.29
501	−0.52	−0.46	+2.66	+2.39	408	−0.21	−0.26
202	−2.42	−2.40	+0.78	+0.96	4, 0, 10	−0.09	−0.07
402	+0.26	+0.07	+0.61	+1.01	608	−6.84	−7.25
103	+3.30	+3.37	−1.57	−1.64	6, 0, 10	−5.57	−5.37
204	−2.87	−3.00	+2.01	+2.11	802	−0.78	−1.00
404	+0.85	+0.40	−0.95	−0.45	8, 0, 10	+1.41	+1.53
105	−1.02	−0.99	−1.83	−2.11	10, 0, 2	+3.53	+3.31
305	−0.17	+0.31	+1.17	+0.91	10, 0, 10	−2.52	−2.39

3. Use of the Charts in Least-Squares Refinement Procedures[2]

If the component of the calculated structure factor $F_{calc}(hk0)$ due to an atom at $(x_1, y_1)=F_1(hk0)$ $=2\cos 2\pi hx_1\cdot\cos 2\pi ky_1$, then $\partial F_{calc}(hk0)/\partial x_1=-4\pi h\cdot\sin 2\pi hx_1\cdot\cos 2\pi ky_1$; $\partial F_{calc}(hk0)/\partial y_1=-4\pi k\cdot\cos 2\pi hx_1\cdot\sin 2\pi ky_1$. Let $F_1'(hk0)=2\cos 2\pi hx_1'\cdot\cos 2\pi ky_1'$ be the component of $F_{obs}(hk0)$ due to the atoms whose true position is (x_1',y_1'). Then

$$F_{obs}(hk0)=\sum_{r=1}^{n} 2\cos 2\pi h(x_r+\Delta x_r)\cdot\cos 2\pi k(y_r+\Delta y_r),$$

and if $\Delta x_r=x_r'-x_r$ and $\Delta y_r=y_r'-y_r$ are small enough so that $\sin 2\pi h\Delta x_r\simeq 2\pi h\Delta x_r$, and $\cos 2\pi h\Delta x_r\simeq 1$, then

$$\begin{aligned}F_{obs}-F_{calc}=\Delta F=-2\pi\{&h[2\sin 2\pi hx_1\cdot\cos 2\pi ky_1\Delta x_1\\&+2\sin 2\pi hx_2\cdot\cos 2\pi ky_2\Delta x_2+\cdots]\\&+k[2\cos 2\pi hx_1\cdot\sin 2\pi ky_1\Delta y_1\\&+2\cos 2\pi hx_2\cdot\sin 2\pi ky_2\Delta y_2+\cdots]\}.\end{aligned}$$

If a least-squares refinement is now carried out in the usual way by forming the normal equations from expressions of this type for the batch of structure factors under consideration, it has been found convenient to discard the factor of 2π until the solutions of the normal equations have been obtained, and to get rid of the minus sign on the right-hand side by using $\Delta F=F_{calc}-F_{obs}$ instead of $F_{obs}-F_{calc}$.

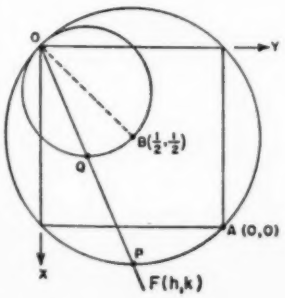


FIG. 8. Demonstration of space-group extinction condition for a face-centered projection.

[2] E. W. Hughes, J. Am. Chem. Soc. 63, 1737 (1941).

Since

$$2\begin{matrix}\cos\\ \sin\end{matrix}2\pi hx_r\cdot\begin{matrix}\sin\\ \cos\end{matrix}2\pi ky_r = \sin 2\pi(hx_r+ky_r)\mp\sin 2\pi(hx_r-ky_r),$$

the coefficients of Δx_r and $\Delta y_r = A_r \mp B_r$ can be obtained by using the (x_r,y_r) circles which have been drawn to calculate the structure factors for the trial structure (Sec. 2 in the foregoing), and recording the readings at the intersections when these circles are superposed on the $\sin 2\pi(hx+ky)$ charts.

In this case the $(hk0)$ coefficient of $\Delta x_1 = {}_1a_{hk0} = h[2\sin 2\pi hx_1\cdot\cos 2\pi ky_1] = h[A_1+B_1]$, and similarly the $(hk0)$ coefficient of $\Delta y_1 = {}_1b_{hk0} = k[2\cos 2\pi hx_1\cdot\sin 2\pi ky_1] = k[A_1-B_1]$. In general, it will be convenient merely to obtain $[A_1+B_1]$ and $[A_1-B_1]$ and multiply by h or k at the same time as multiplying by the atomic scattering factor $f(hk0)$.

As an example of the accuracy to be expected when ${}_ra_{hk0}$ and ${}_rb_{hk0}$ are obtained graphically from the standard charts, the shifts were computed for x_1 and y_1, again using the example referred to in Sec. 2 for comparison with the values obtained when ${}_ra_{hk0}$ and ${}_rb_{hk0}$ and $\Delta F(hk0)$ were evaluated to the nearest 0.001; all the equations were given equal weight, as would be possible in calculations involving monatomic structures when the atomic scattering factor can be taken out of the equations by multiplying $F_{obs}(hk0)$ by $1/f(hk0)$. After forming $({}_ra_{hk0})^2$ and ${}_ra_{hk0}\cdot\Delta F(hk0)$ and proceeding in the usual way, but using only the diagonal terms, the following results were obtained:

$$\Delta x_1(\text{calc}) = -0.010 \qquad \Delta y_1(\text{calc}) = -0.005$$
$$\Delta x_1(\text{graph}) = -0.008 \qquad \Delta y_1(\text{graph}) = -0.004.$$

These results are felt to be very satisfactory.

As in the case of structure factor calculations, use of the charts for the application of least squares to structure factors of the form $F(hk0)=\sum_r \cos 2\pi(hx_r+ky_r)$ is particularly convenient, since

$$\frac{\partial F(hk0)}{\partial x_1} = -2\pi h[\sin 2\pi(hx_r+ky_r)]$$

and

$$\frac{\partial F(hk0)}{\partial y_1} = -2\pi k[\sin 2\pi(hx_r+ky_r)]$$

so that both the Δx_r and the Δy_r coefficients are obtained from one single intersection on the $\sin 2\pi(hx+ky)$ chart, the factor $-2\pi h$ or $-2\pi k$ being taken care of in the same way as before.

4. Improvement of Trial Structures by Visual Estimation of Errors

By means of this construction it is possible to carry out a "trial and error" type of refinement. Suppose that in the case where $F(hk0)=\sum_r \cos 2\pi(hx_r+ky_r)$, a set of circles has been drawn for the atomic positions in a trial structure, and used to obtain $F_{calc}(hk0)$ values, and hence $\Delta F(hk0)$ values. If these circles are superposed on a $\cos 2\pi(hx+ky)$ chart, it can be seen by inspection whether $F(hk0)$ will increase or decrease for an increase of x or y for any atom, and whether it will vary slowly or rapidly. Thus a movement of one or several circles can be sought which will produce changes in $F(hk0)$ in the right direction. The method has the advantage in that the effect of any change in the atomic position can be readily observed for a large number of reflections simultaneously.

If attention is directed to successive groups of reflections for which the interplanar spacing, and hence $(\sin\theta)/\lambda$, varies only over a small range, the relative intensities of the reflections within such a group will still be significant even before a temperature factor has been assigned, and the set of ratios $\sum|F_{calc}|/\sum|F_{obs}|$ for the series of groups can then be used to evaluate it, especially in the case of monatomic crystals.

The Necking of a Tension Specimen in Plane Plastic Flow*

E. T. ONAT† AND W. PRAGER‡
Brown University, Providence, Rhode Island
(Received September 28, 1953)

A linearized version of the equations of plasticity is used to discuss the symmetrical necking of a tension specimen in plane plastic flow. The material is treated as perfectly plastic, and elastic strains are neglected.

1. INTRODUCTION

THE complete necking process of certain deeply notched tension specimens in plane plastic flow has been discussed by Lee[1,2] and Wang.[3] The necking of a specimen with deep notches of arbitrary shape can be treated by the graphical procedure of Prager.[4] The necking of a seemingly prismatic specimen, however, may be caused by slight initial imperfections in the form of long and extremely shallow grooves. In this case a linearization of the equations of plasticity is justified that simplifies the analytical treatment of the necking process.

Linearization of the differential equations or the boundary conditions has long been used successfully in fluid dynamics and has yielded valuable approximate solutions to otherwise intractable problems. It is therefore surprising that this technique has not been used as widely in the mathematical theory of plasticity. In certain problems of rotationally symmetric plastic flow, the substitution of Tresca's yield condition for that of v. Mises may, of course, be regarded as a linearization of the yield condition. A somewhat more deliberate use of linearization is found in a recent paper by Gaydon,[5] but this author linearizes only the stress equations and not the velocity equations. The example treated in the present paper would seem to suggest that joint linearization of stress and velocity equations greatly facilitates the treatment of certain problems of plastic flow.

2. THE LINEARIZED EQUATIONS

Using rectangular coordinates, x, y, z with the x axis in the direction of the applied tension and the z axis normal to the plane of flow, denote the relevant stress components in the usual manner by σ_x, σ_y, σ_z, and $\tau_{xy}=\tau$, and the velocity components by u and v. Those components of the strain rate that do not vanish identically are

$$\epsilon_x=\frac{\partial u}{\partial x},\quad \epsilon_y=\frac{\partial v}{\partial y},\quad \gamma_{xy}=\gamma=\frac{\partial u}{\partial y}+\frac{\partial v}{\partial x}. \tag{1}$$

If the material obeys the yield condition and flow rule of v. Mises and is in plane plastic flow,[6] then

$$\sigma_z=\tfrac{1}{2}(\sigma_x+\sigma_y). \tag{2}$$

Furthermore, in the absence of body forces and for quasistatic flow, the stresses σ_x, σ_y, τ must satisfy the equations of equilibrium,

$$\frac{\partial\sigma_x}{\partial x}+\frac{\partial\tau}{\partial y}=0,\quad \frac{\partial\sigma_y}{\partial y}+\frac{\partial\tau}{\partial x}=0, \tag{3}$$

and the yield condition,

$$(\sigma_x-\sigma_y)^2+4\tau^2=4k^2, \tag{4}$$

where k is the yield limit in simple shear. Finally, the flow rule of v. Mises demands that[6]

$$\epsilon_x-\epsilon_y=0,\quad \frac{\epsilon_x-\epsilon_y}{\gamma}=\frac{\sigma_x-\sigma_y}{2\tau}, \tag{5}$$

where $sgn\gamma=sgn\tau$.

Equations (1), (3), (4), and (5) constitute a system of eight differential equations for the unknowns σ_x, σ_y, τ, ϵ_x, ϵ_y, γ, u, and v. With the exception of (4) and the second Eq. (5), these equations are linear in the unknown quantities.

For the necking problem considered in the following, these equations are linearized by assuming that the state of stress throughout the plastic portion of the specimen does not differ much from the basic state of stress specified by $\sigma_x=2k$, $\sigma_y=\tau=0$. When the left-hand side of (4) is replaced by its Taylor development in the neighborhood of the basic state of stress and higher-order terms are neglected, the linearized yield condition

$$\sigma_x-\sigma_y=2k \tag{6}$$

is obtained. If now this linearized yield condition is used as the plastic potential (reference 6, Chapter III,

* The results presented in this paper were obtained in the course of research conducted under Contract DA-19-020-ORD-2598. The authors are indebted to the authorities of Watertown Arsenal for the permission to publish these results.
† F. B. Jewett Fellow in Applied Mathematics, Brown University.
‡ Professor of Applied Mechanics, Brown University.
[1] E. H. Lee, J. Appl. Mech. 19, 331–336 (1952).
[2] E. H. Lee, J. Appl. Mech. (to be published).
[3] A. J. Wang, Quart. Appl. Math. 11, 427–438 (1954).
[4] W. Prager, Trans. Roy. Inst. Tech. (Stockholm, Sweden), No. 65, 1953.
[5] F. A. Gaydon, J. Mech. Phys. Solids 1, 103–112 (1953).
[6] R. Hill, *The Mathematical Theory of Plasticity* (Clarendon Press, Oxford, 1950), Chapter VI, Section 1.

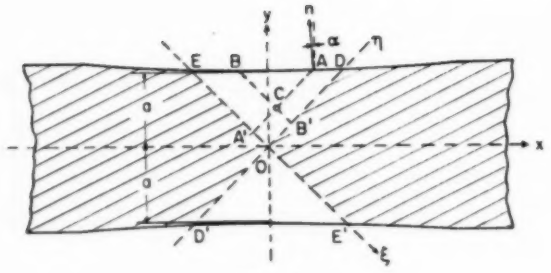


FIG. 1.

Section 1), the linearized flow rule

$$\epsilon_x+\epsilon_y=0, \quad \gamma=0 \tag{7}$$

is obtained instead of (5).

By the use of (6), the stress σ_y can be eliminated from the equations of equilibrium (3). Thus,

$$\frac{\partial\sigma_x}{\partial x}+\frac{\partial\tau}{\partial y}=0, \quad \frac{\partial\sigma_x}{\partial y}+\frac{\partial\tau}{\partial x}=0. \tag{8}$$

This system of partial differential equations for σ_x and τ is hyperbolic. With the characteristic variables

$$\xi=x-y, \quad \eta=x+y, \tag{9}$$

Eqs. (8) can be written in the form

$$\frac{\partial}{\partial\eta}(\sigma_x+\tau)=0, \quad \frac{\partial}{\partial\xi}(\sigma_x-\tau)=0. \tag{10}$$

It follows from the linearized yield condition (6) that Eqs. (10) remain valid when σ_x is replaced by σ_y.

Substitution of (1) into the linearized flow rule (7) yields

$$\frac{\partial u}{\partial x}+\frac{\partial v}{\partial y}=0, \quad \frac{\partial u}{\partial y}+\frac{\partial v}{\partial x}=0. \tag{11}$$

These equations have the same form as Eqs. (8) with u and v replacing σ_x and τ, respectively. Thus, in analogy to (10), we have

$$\frac{\partial}{\partial\eta}(u+v)=0, \quad \frac{\partial}{\partial\xi}(u-v)=0. \tag{12}$$

3. SYMMETRICAL NECKING

For simplicity, consider a tension specimen that is weakened by two symmetrical shallow grooves. Choosing the axes of x and y as is indicated in Fig. 1, let the shape of the grooves be specified by

$$y=\pm(a+cx^2), \tag{13}$$

where $2a$ is the smallest width of the specimen and c is a constant. In the following, ac will be treated as a small quantity and terms of the second or higher orders in this quantity will be neglected.

Let A be a generic point on the contour of the upper groove in Fig. 1, n the exterior normal to this contour at A and α the angle between the positive y axis and n. From (13), the small angle α is seen to be given by

$$\alpha=dy/dx=2cx. \tag{14}$$

Since the surface tractions at A vanish, we have

$$\begin{aligned}\sigma_x(1-\cos2\alpha)+\sigma_y(1+\cos2\alpha)-2\tau\sin2\alpha=0,\\ (\sigma_x-\sigma_y)\sin2\alpha-2\tau\cos2\alpha=0.\end{aligned} \tag{15}$$

With the approximations $\cos2\alpha\approx1$ and $\sin2\alpha\approx4cx$, Eqs. (6) and (15) furnish the following stresses at A:

$$\sigma_{xA}=2k, \quad \sigma_{yA}=0, \quad \tau_A=4ckx_A, \tag{16}$$

where the subscript A refers to the point A.

To obtain the stresses at an interior point C, draw the characteristics through this point and label their intersections with the groove contour A and B as in Fig. 1. According to (10), the sum $\sigma_x+\tau$ has a constant value along AC, and the difference $\sigma_x-\tau$ has a constant value along BC. Thus, using (16), we find

$$\begin{aligned}\sigma_{xC}&=\tfrac{1}{2}(\sigma_{xA}+\sigma_{xB}+\tau_A-\tau_B)=2k[1+c(x_A-x_B)],\\ \tau_C&=\tfrac{1}{2}(\sigma_{xA}-\sigma_{xB}+\tau_A+\tau_B)=2ck(x_A+x_B).\end{aligned} \tag{17}$$

The yield condition (6) then furnishes

$$\sigma_{yC}=2ck(x_A-x_B). \tag{18}$$

Within the framework of our approximation, $x_A-x_B=2(a-y_C)$ and $x_A+x_B=2x_C$. Substituting these expressions into Eqs. (17) and (18), and omitting the subscript C, we obtain the stresses in the upper plastic region in Fig. 1 as follows:

$$\sigma_x=2k[1+2c(a-y)], \quad \sigma_y=4ck(a-y), \quad \tau=4ckx. \tag{19}$$

From the first Eq. (19), the axial force acting on the specimen is found to be

$$F=2\int_0^a\sigma_x dy=4ak(1+ac). \tag{20}$$

During the necking process considered here, the portions of the specimen that are shaded in Fig. 1 remain rigid and move with unit speed parallel to the x axis and away from the origin. Along OD we have therefore $u=1$, $v=0$ on the shaded side of OD. Similarly, along OE we have $u=-1$, $v=0$ on the shaded side of OE. Since only the velocity components normal to the lines OD and OE need to be continuous across these lines, all that can be stated regarding the velocities on the unshaded sides of these lines is that $u-v=1$ along OD and $u+v=-1$ along OE. These boundary conditions enable us to determine the velocity field in the region DOE. To find the velocity components at an interior point C of this region, draw the characteristics through this point and label their intersections with the domain boundary A' and B' as in Fig. 1. According to

(12) and the boundary conditions established above, $u+v$ has the constant value -1 along $A'C$, and $u-v$ has the constant value 1 along $B'C$. At C, we have therefore

$$u=0, \quad v=-1. \tag{21}$$

Equations (21) show that the entire region DOE moves as a rigid body with unit speed in the negative y direction. Similarly, $D'OE'$ moves as a rigid body with unit speed in the positive y direction, and the instantaneous plastic flow is restricted to the lines of discontinuity DD' and EE'.

On account of the great simplicity of this velocity field, finite plastic deformation is readily discussed. Figure 2 shows the specimen in a later stage. Plastic deformation has occurred only in the shaded portions; the unshaded portions have moved as rigid bodies. At the considered instant, all points to the right of $DD^*O^*E^{*\prime}E$ move with unit speed in the position x direction, all points above $E^*O^*D^*$ move with unit speed in the negative y direction, etc., the instantaneous plastic flow being again restricted to the lines of discontinuity $D^*D^{*\prime}$ and $E^*E^{*\prime}$. Since the arc D^*E^* in Fig. 2 is congruent with the central part of the arc DE in Fig. 1, the stresses in the region $D^*O^*E^*$ are given by the equations obtained from (19) by substituting a^* for a. If the considered necking process continues without rupture, the central portion of the specimen will eventually consist of two wedges with common apex. If the initial grooves were rather shallow, these wedges will have practically rectilinear sides that have the slope 1:2 with respect to the x axis.

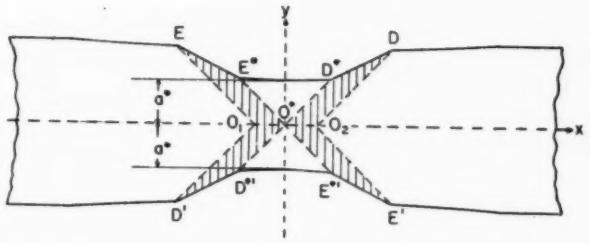


FIG. 2.

It is well known (see, for instance, reference 6, p. 251) that the plastic stress and velocity fields considered here are possible only if the initial grooves are "sufficiently deep." The precise criterion for "sufficient depth" is not known. For the parabolic groove specified by (13), the usual estimate of the necessary depth leads to the statement that the length of the groove should be at least $4a$. This condition is fulfilled in Fig. 1.

JOURNAL OF APPLIED PHYSICS VOLUME 25, NUMBER 4 APRIL, 1954

The Growth of Vapor Bubbles in Superheated Liquids*

M. S. PLESSET AND S. A. ZWICK
California Institute of Technology, Pasadena, California

(Received September 26, 1953)

The growth of a vapor bubble in a superheated liquid is controlled by three factors: the inertia of the liquid, the surface tension, and the vapor pressure. As the bubble grows, evaporation takes place at the bubble boundary, and the temperature and vapor pressure in the bubble are thereby decreased. The heat inflow requirement of evaporation, however, depends on the rate of bubble growth, so that the dynamic problem is linked with a heat diffusion problem. Since the heat diffusion problem has been solved, a quantitative formulation of the dynamic problem can be given. A solution for the radius of the vapor bubble as a function of time is obtained which is valid for sufficiently large radius. This asymptotic solution covers the range of physical interest since the radius at which it becomes valid is near the lower limit of experimental observation. It shows the strong effect of heat diffusion on the rate of bubble growth. Comparison of the predicted radius-time behavior is made with experimental observations in superheated water, and very good agreement is found.

I. INTRODUCTION

WHEN the vapor pressure in a liquid exceeds the ambient pressure, it becomes possible for a vapor bubble to grow from a small "nucleus" in the liquid. This nucleus is a region of nonliquid phase and presumably consists of a gas or vapor phase stabilized on a solid particle. The rate of growth of a vapor bubble, once formed, is determined by the surface tension, the liquid inertia, and the difference between the pressure within the bubble and the ambient, or external, pressure. In the initial stages of the bubble expansion, when the forces are nearly in equilibrium, the growth is slow but it is accelerated with increase in bubble size because of the reduction in surface tension. When the rate of bubble growth becomes appreciable, however, the temperature and, hence, the pressure within the bubble drop and the rate of growth is decreased. One might, therefore, expect a maximum in the velocity of the bubble wall. The reduction of the temperature

* This study was supported by the U. S. Office of Naval Research.

within the bubble is a consequence of the latent heat requirement of the evaporation which takes place at the vapor-liquid interface as the bubble grows.

For the quantitative solution of the problem, some simplifying physical assumptions may be made. The example considered in detail here is that of a vapor bubble growing in moderately superheated water, and the arguments for the validity of the general assumptions are justified for this case.

It will be assumed that the bubble is spherical throughout its growth. This assumption is reasonable provided the radial acceleration and velocity are small, for with a spherically symmetric external pressure field the spherical shape will then be stable under the action of surface tension.[1] Excluded from consideration is the asymmetric, buoyant force of gravity which becomes important if the bubble growth is followed for so long a time that a significant translational velocity is acquired. A translational velocity of the bubble as a whole not only causes a deformation in shape but also increases the rate of heat inflow to the bubble over that used in the analysis presented below. In water superheated by about 10°C, no great error is introduced by the buoyant force provided the bubble growth is not followed beyond a radius of approximately 1 mm.

For a superheat of 10°C in water, a vapor bubble grows from its initial microscopic size to a radius of 1 mm in a time of the order of 10 msec. The corresponding average radial velocity of 10 cm/sec is very small compared with the velocity of sound in the liquid, so that compressibility effects may be entirely neglected in the motion of the liquid produced by the bubble growth. It may also be shown that viscous effects are unimportant.[2] The hydrodynamic equation of motion, which contains the effect of the liquid inertia, is thus greatly simplified. The bubble wall velocity is also very small compared with the velocity of sound in the vapor so that compressibility effects are unimportant here as well. One has the further simplification in the vapor region that the vapor density is so small that its inertia effects may be neglected. It then follows from Bernoulli's equation applied to a fluid of negligible density that the pressure within the vapor region may be taken as uniform. In addition, since the acoustic velocity in the vapor is so large, the pressure within the vapor follows practically instantaneously its value at the bubble wall. When the velocity of the bubble wall is sufficiently slow, the pressure of the vapor is given by the equilibrium vapor pressure of the liquid. That this is the case for the present problem may be seen as follows. The mean velocity appropriate for the rate of evaporation from a liquid surface is $(BT/2\pi M)^{\frac{1}{2}}$, where B is the gas constant, T the absolute temperature, and M is the molecular weight. This velocity must be reduced by the coefficient for evaporation which has the value 0.04 for water.[3] The critical velocity for a water surface at approximately 100°C is therefore about 8 m/sec, which is appreciably greater than the radial velocities encountered here, so that the pressure deficiency from the equilibrium vapor pressure may be neglected.

The temperature of the vapor would, in general, vary with position in the bubble as well as with time. The approximation will be made, however, that the thermal diffusivity D' of the vapor is so large that temperature gradients are negligible within the bubble. In water vapor, the characteristic diffusion length $(2D't)^{\frac{1}{2}}$ is 0.24 cm for $t=10^{-2}$ sec. For the greatest superheat considered here (≈ 10°C) the bubble radius at this time is approximately 0.1 cm, which is significantly smaller than the characteristic diffusion length. The approximation of uniform temperature within the vapor improves with decrease in superheat.

In summary, the physical model upon which the calculations are based consists of a spherical vapor bubble which has uniform temperature and pressure; the temperature of the vapor is that of the liquid at the bubble wall, and the pressure is the equilibrium vapor pressure for that temperature. In addition, the effects of viscosity and compressibility are neglected both in the vapor and in the liquid.

The equation of motion for the radius R of the bubble in a nonviscous, incompressible liquid as a function of time is[4]

$$R(d^2R/dt^2)+\tfrac{3}{2}\dot{R}^2=[p(R)-P_0]/\rho, \qquad (1)$$

where ρ is the liquid density, P_0 is the external pressure in the liquid or the pressure at infinity, and $p(R)$ is the pressure in the liquid at the bubble boundary. A dot denotes differentiation with respect to time. The pressure $p(R)$ is given in terms of the vapor pressure p_v within the bubble by

$$p(R)=p_v(T)-2\sigma/R, \qquad (2)$$

where σ is the surface tension constant and $p_v(T)$ is the equilibrium vapor pressure for the temperature T at the bubble boundary. In the following, the small variations of σ and ρ with temperature are neglected. It is convenient to introduce a radius R_0 defined by the relation

$$2\sigma/R_0=p_v(T_0)-P_0, \qquad (3)$$

in which T_0 is the temperature of superheat of the liquid at a distance from the bubble. Physically, R_0 is the effective initial radius of the bubble; it represents an extrapolation of the free spherical bubble down to the equilibrium radius for the given initial conditions. It should be noted that a bubble at rest with radius R_0 is in unstable equilibrium. The actual nucleus from which the bubble grows is not necessarily spherical and its surface energy may be appreciably less than $4\pi\sigma R_0^2$;

[1] M. S. Plesset, J. Appl. Phys. **25**, 96 (1954).
[2] F. R. Gilmore, Report 26-4, Hydrodynamics Laboratory, California Institute of Technology, April 1952.
[3] G. Wyllie, Proc. Roy. Soc. (London) **A197**, 383 (1949).
[4] G. Lamb, *Hydrodynamics* (Dover Publications, New York, 1945); M. S. Plesset, J. Appl. Mechanics **16**, 277 (1949).

however, the nucleus from which the bubble grows and the free spherical bubble of radius R_0 are both in unstable equilibrium at the temperature T_0 and external pressure P_0. Table I gives a set of values of R_0 at various superheat temperatures in water for a pressure P_0 of one atmos.

In terms of the parameter R_0, Eq. (1) may be written

$$R(d^2R/dt^2)+\tfrac{3}{2}\dot{R}^2 = \frac{p_v(T)-p_v(T_0)+(2\sigma/R_0)(1-R_0/R)}{\rho}, \quad (4)$$

or

$$\frac{1}{2R^2\dot{R}}\frac{d}{dt}(R^3\dot{R}^2) = \frac{p_v(T)-p_v(T_0)+(2\sigma/R_0)(1-R_0/R)}{\rho}. \quad (5)$$

If the cooling effect of evaporation is disregarded, so that $p_v(T)=p_v(T_0)$, Eq. (5) may be integrated to give

$$\dot{R}^2=(R_0/R)\dot{R}_0^2+(4\sigma/3\rho R_0)[1-R_0^3/R^3] - (2\sigma/\rho R)[1-R_0^2/R^2]. \quad (6)$$

This solution will be referred to as the Rayleigh solution.[5] For $R\gg R_0$, Eq. (6) becomes

$$\dot{R}^2\sim\frac{4\sigma}{3\rho R_0}=\frac{2}{3\rho}[p_v(T_0)-P_0], \quad (7)$$

which is a constant.

The actual motion deviates markedly from that predicted by the Rayleigh solution because of the cooling effect. The heat $\dot{Q}$ which must be supplied to the bubble per unit time is

$$\dot{Q}=(4\pi/3)L(d/dt)(R^3\rho'), \quad (8)$$

where L is the latent heat of evaporation per unit mass and ρ' is the vapor density. This heat is supplied by conduction from the liquid into the bubble so that

$$\dot{Q}=4\pi R^2k(\partial T/\partial r)_R, \quad (9)$$

where k is the thermal conductivity of the liquid and $(\partial T/\partial r)_R$ is the temperature gradient in the liquid at the bubble boundary. Hence,

$$(\partial T/\partial r)_R=(L/3k)(1/R^2)(d/dt)(R^3\rho'). \quad (10)$$

The small variations of L and k with temperature are neglected. It may be shown that the contribution of the term proportional to $d\rho'/dt$ is much less than that from the term proportional to dR^3/dt. Equation (10) may be approximated, therefore, by

$$(\partial T/\partial r)_R\approx(L\rho'/k)\dot{R}. \quad (11)$$

[5] The solution for the motion of a bubble under such constant pressure conditions was given by Rayleigh, Phil. Mag. 34, 94 (1917) and applied by him to the case of a collapsing bubble.

TABLE I. Effective nucleus size R_0 in water at one atmosphere for the values of the temperature T_0.

R_0 (cm×10³)	1.56	1.02	0.751	0.590	0.483
T_0 (°C)	102	103	104	105	106

This approximation is plausible physically since the volume increases by orders of magnitude, while the vapor density changes very little because of the small temperature decrease during the bubble growth.

Equation (11) together with the specification of the temperature at infinity T_0 determines the temperature field in the liquid. The solution for this temperature problem with the moving boundary $R(t)$ has been found[6] under the assumption that the drop in temperature from T_0 to the value T at the bubble boundary takes place in a layer of liquid adjacent to the bubble which has small thickness compared with $R(t)$. This approximation of the "thin thermal boundary layer" is justified physically because the thermal diffusivity of the liquid is small. The approximate expression for the temperature at the bubble wall is[7]

$$T=T_0-\left(\frac{D}{\pi}\right)^{\frac{1}{2}}\int_0^t\frac{R^2(x)(\partial T/\partial r)_{r=R(x)}}{\left\{\int_x^t R^4(y)dy\right\}^{\frac{1}{2}}}dx. \quad (12)$$

The temperature T is thus given in terms of R and $\dot{R}$ by Eq. (12) so that $p_v(T)$ is also specified in terms of these variables. Equations (5) and (12) therefore determine the dynamic problem of the bubble growth.

II. SOLUTION OF THE EQUATION OF MOTION

Equations (5) and (12) are connected when the equilibrium vapor pressure is specified as a function of temperature. For superheats not too far above the boiling temperature T_b of the liquid at the external pressure P_0, the vapor pressure may be approximated by a linear function of the temperature,

$$\frac{p_v(T)-P_0}{\rho}=A(T-T_b). \quad (13)$$

For $T=T_0$, Eq. (13) gives

$$2\sigma/\rho R_0=A(T_0-T_b) \quad (14)$$

because of Eq. (3). Equation (5) may now be written as

$$\frac{1}{2R^2\dot{R}}\frac{d}{dt}(R^3\dot{R}^2)=A(T-T_0)+\frac{2\sigma}{\rho R_0}\left(1-\frac{R}{R_0}\right). \quad (15)$$

[6] M. S. Plesset and S. A. Zwick, J. Appl. Phys. 23, 95 (1952).

[7] Reference 6, Eq. (20). The error in Eq. (20) may be estimated from Eq. (31) of this reference. For the problem discussed here it was found to be less than 10 percent of the difference $T-T_0$ at any time.

It is convenient to use in place of R the dimensionless volume ratio

$$p=R^3/R_0^3, \tag{16a}$$

and in place of t the dimensionless variable

$$u=(\alpha/R_0^4)\int_0^t R^4(y)dy, \tag{16b}$$

where the constant α is defined as

$$\alpha=(2\sigma/\rho R_0^3)^{\frac{1}{2}} \tag{16c}$$

and has the dimensions of reciprocal time. Equation (15) then is transformed to

$$\frac{1}{6p'}\frac{d}{du}(p^{7/3}p'^2)=1-\frac{1}{p^{\frac{1}{2}}}-\mu\int_0^u\frac{p'(v)dv}{(u-v)^{\frac{1}{2}}}, \tag{17}$$

in which $p'=dp/du$ and the dimensionless parameter μ is given by

$$\mu=\frac{AL\rho'}{3kR_0\alpha}\left(\frac{D}{\pi\alpha}\right)^{\frac{1}{2}}. \tag{18}$$

The physical quantities of interest are now expressed as follows:

$$t=\frac{1}{\alpha}\int_0^u\frac{dv}{p^{4/3}(v)}; \tag{19a}$$

$$R=R_0p^{\frac{1}{3}}; \tag{19b}$$

$$R=(\alpha R_0/3)p^{\frac{1}{3}}p'; \tag{19c}$$

$$T-T_0=-\frac{\alpha^2R_0^2\mu}{A}\int_0^u\frac{p'(v)dv}{(u-v)^{\frac{1}{2}}}. \tag{19d}$$

The values of the radius R for which experimental observations are readily obtainable are much greater than R_0 so that the asymptotic solution of Eq. (17) ($p\gg1$) is of the greatest physical interest. It is of importance, however, to examine the initial stages of the bubble growth since a qualitative description of this portion of the problem is necessary for an understanding of the asymptotic solution. It will be shown that the solution in the asymptotic range is not affected by the details of the mathematical model used to describe the behavior of the bubble for R near R_0 so that the uncertainty in the surface energy for small values of R causes no difficulties in the physical applications to be made here.

A. Initial Growth of the Bubble

It is evident from Eq. (17) that a bubble of radius $R_0(p=1)$ initially at rest ($p'(0)=0$) will remain at rest in unstable equilibrium. The present mathematical model requires some change in the external conditions before bubble growth will be initiated.

The particular mechanism which will be considered here is a rise in the temperature of the liquid. Suppose, for example, that there is a constant heat source of strength a per unit volume so that the amount of heat added per unit volume η from $t=0$ to $t=t$ is

$$\eta=at. \tag{20}$$

Then, if the temperature of the bulk liquid is T_0 at $t=0$, at time t it will be

$$T_\infty=T_0+\eta/\rho c=T_0+\eta D/k. \tag{21}$$

The heat flow problem may be solved[6] for the boundary conditions of Eqs. (21) and (10), and the temperature at the bubble wall is given by replacing T_0 in Eq. (12) by T_∞. In place of Eq. (19d) one has

$$T=T_0+\eta D/k-\frac{\alpha^2R_0^2\mu}{A}\int_0^u\frac{p'(v)dv}{(u-v)^{\frac{1}{2}}}. \tag{22}$$

The extension of Eq. (17) to include the heat source term is thus

$$\frac{1}{6p'}\frac{d}{du}(p^{7/3}p'^2)=1-\frac{1}{p^{\frac{1}{2}}}-\mu\int_0^u\frac{p'(v)dv}{(u-v)^{\frac{1}{2}}}+\frac{AD}{\alpha^2R_0^2k}\eta. \tag{23a}$$

For values of p near the initial value of unity, one may write

$$\eta=at=\frac{a}{\alpha}\int_0^u\frac{dv}{p^{4/3}(v)}\approx\frac{a}{\alpha}u,$$

where the substitution for t follows from Eq. (19a). Equation (23a) now becomes

$$\frac{1}{6p'}\frac{d}{du}(p^{7/3}p'^2)=1-\frac{1}{p^{\frac{1}{2}}}-\mu\int_0^u\frac{p'(v)dv}{(u-v)^{\frac{1}{2}}}+\gamma u, \tag{23b}$$

where

$$\gamma=ADa/\alpha^3R_0^2k. \tag{24}$$

The constant γ is extremely small; for example, in water with a temperature rise of 1°C/sec and $R_0=10^{-3}$ cm, γ is approximately 10^{-6}. The smallness of the constant γ implies that the forced growth away from the unstable equilibrium point $p=1$, $p'=0$ is very slow until the bubble radius has increased sufficiently for the surface tension to be partially "relaxed." This initial slow growth is a delay period in the bubble growth, for the bubble radius changes very little until the remaining terms on the right side of Eq. (23b) become appreciable. The bubble growth is then so rapid that the change in the bulk temperature of the liquid is insignificant, and the term in γ may be neglected. The delay period effectively gives an initial value of R slightly greater than R_0 from which the important growth begins. An approximate solution may be found for this initial period of forced growth from the equilibrium point by linearizing Eq. (23b); that is, by neglecting terms of the second order in $p-1$ and its derivatives. The details of this

TABLE II. Values of the parameters for water at various superheat temperatures. The external pressure $P_0=1$ atmos; the constant A has been given the value 4.08×10^4 dynes cm/(°C g).

T_0	102°	103°	104°	105°	106°
R_0	1.558×10^{-3}	1.019×10^{-3}	7.507×10^{-4}	5.901×10^{-4}	4.832×10^{-4}
α	1.797×10^{5}	3.391×10^{5}	5.356×10^{5}	7.677×10^{5}	1.035×10^{6}
μ	5.598×10^{-1}	3.411×10^{-1}	2.407×10^{-1}	1.841×10^{-1}	1.480×10^{-1}
b_1	-1.073	-9.099×10^{-1}	-8.101×10^{-1}	-7.409×10^{-1}	-6.890×10^{-1}
b_2	-4.709×10^{-1}	-4.624×10^{-1}	-7.122×10^{-1}	-1.334	-2.506
b_3	2.339×10^{-1}	-1.481×10^{-3}	-3.586×10^{-1}	-9.725×10^{-1}	-1.972
b_4	-5.534×10^{-1}	1.258×10^{-4}	-1.970×10^{-1}	-4.598	-2.510×10^{1}
b_5	-1.775	3.064×10^{-1}	4.298	1.850×10^{1}	6.584×10^{1}
b_6	-7.423×10^{-1}	-2.670×10^{-1}	6.166×10^{-1}	-4.736×10^{-1}	-4.865×10^{1}

calculation will not be presented here; but, as would be expected, the forced growth away from the equilibrium has an exponential behavior,

$$R\approx R_0\left\{1+\frac{2\gamma}{\beta(3\beta^2+1)}e^{\alpha\beta t}\right\}. \tag{25}$$

The temperature at the bubble wall is given approximately by

$$T\approx T_\infty-\frac{R_0^2\alpha^2}{A}\frac{2\gamma(1-\beta^2)}{\beta(3\beta^2+1)}e^{\alpha\beta t}. \tag{26}$$

The constant β is that root of the equation

$$\beta^2+3\mu(\pi\beta)^{\frac{1}{2}}-1=0$$

for which $\beta^{\frac{1}{2}}$ is positive.

B. Growth Behavior for $R\gg R_0$

After the bubble growth has been initiated, there is a rapid rise in the velocity $\dot{R}$ until the cooling effect becomes important. The bubble wall velocity decreases continuously thereafter. No observations on bubble growth in this region are as yet available and the details of this analysis will be omitted. Of present interest is the asymptotic phase of the bubble growth determined by Eq. (17) which is characterized by the limiting effect on $\dot{R}$ of the heat diffusion from the liquid to the vapor. As R increases, the temperature at the bubble wall decreases steadily, but it cannot fall below T_b; for, if the vapor temperature fell below T_b, the pressure difference p_v-P_0 would become negative and the bubble growth would be arrested and eventually reversed. Such behavior is excluded on physical grounds. It therefore follows that the integral on the right-hand side of Eq. (17), which is proportional to the temperature drop at the bubble wall, must approach a limit as t or $u\to\infty$. A further physical argument determines more precisely the asymptotic behavior of this integral. The left-hand side of Eq. (17) represents essentially the acceleration effects of the bubble growth in the liquid. As the bubble grows, this acceleration tends toward zero because of the cooling effect. It therefore follows that

$$\mu\int_0^u\frac{p'(v)dv}{(u-v)^{\frac{1}{2}}}\sim1 \quad \text{as} \quad u\to\infty. \tag{27}$$

This asymptotic relation may be inverted to yield[8]

$$p(u)\approx(2/\pi\mu)u^{\frac{1}{2}} \quad \text{as} \quad u\to\infty. \tag{28}$$

If this result is substituted in Eq. (17), it may be verified that Eqs. (17), (27), and (28) are, in fact, consistent.

Equation (28) is not yet useful since it provides no means of matching the indicated asymptotic solution of Eq. (17) with a solution valid for small values of p. The possibility of matching solutions depends on the possibility of shifting the asymptotic solution in t (or in u) so as to account for the delay period in bubble growth. It is necessary that one be free to shift the asymptotic solution since the duration of the delay period depends completely on the choice of the heat source term while the subsequent behavior of the bubble is independent of this term. A means for making an arbitrary time shift is furnished by noting that, in addition to the asymptotic solution of Eq. (28), Eq. (17) also possesses the solution $p(u)=1$. It will therefore be assumed that the asymptotic solution is described by

$$p(u)=1, \qquad 0<u\leq u_1, \tag{29}$$

$$\mu\int_{u_1}^u\frac{p'(v)dv}{(u-v)^{\frac{1}{2}}}=1-\frac{1}{p^{\frac{1}{2}}}-\frac{1}{6p'}\frac{d}{du}(p^{7/3}p'^{2}), \quad u>u_1. \tag{30}$$

From Eq. (19a), one has corresponding to Eqs. (29) and (30)

$$t=\frac{1}{\alpha}\int_0^u\frac{dv}{p^{4/3}(v)}=\begin{cases}\dfrac{u}{\alpha}, & 0<u\leq u_1,\\[2ex] \dfrac{u_1}{\alpha}+\dfrac{1}{\alpha}\displaystyle\int_{u_1}^u\frac{dv}{p^{4/3}}, & u>u_1,\end{cases} \tag{31}$$

so that u_1/α represents the duration of the delay period in the growth. The time delay may be introduced explicitly in the asymptotic solution for $u>u_1$ by use of the fact that, if $p(u)$ is a solution of Eq. (30), $p(u+u_0)$ is also a solution where u_0 is a constant. A consistent scheme for continuing the asymptotic solution may then

[8] Equation (27) is multiplied by $(z-u)^{-\frac{1}{2}}$ and integrated from $u=0$ to $u=z$ to give Eq. (28).

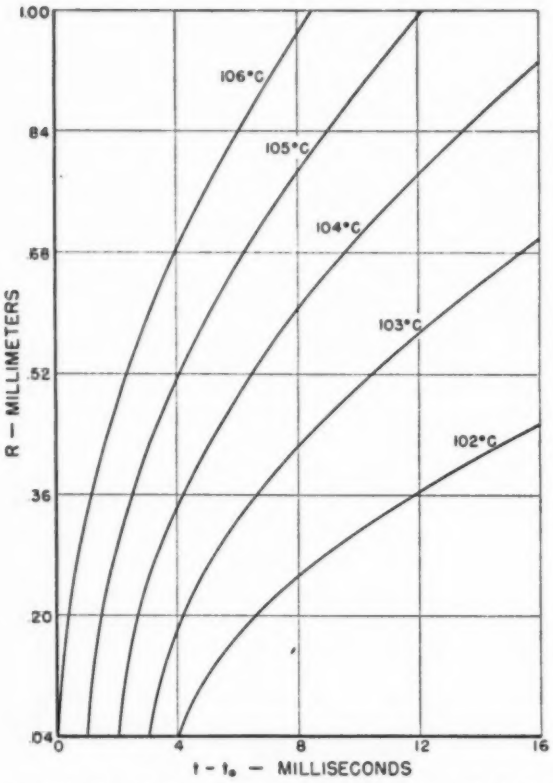


FIG. 1. Calculated bubble radius *versus* time curves at the indicated superheat temperatures in water at 1 atmos.

be found by taking the solution to be of the form

$$p(u)=1, \quad 0<u\leq u_1, \tag{32a}$$

$$p(u)\sim\frac{2}{\pi\mu}(u-u_0)^{\frac{1}{2}}\left\{1+\frac{b_1}{(u-u_0)^{1/6}}+\cdots+\frac{b_5}{(u-u_0)^{5/6}}+\frac{b_6\ln(u-u_0)}{(u-u_0)}\right\}, \quad u>u_1 \tag{32b}$$

to seven terms.[9] When the coefficients b_k have been determined, the difference (u_1-u_0) is fixed by the requirement that $p(u_1)=1$. The delay period is then determined by the choice of u_0.

When (32b) is substituted in the integral on the left side of Eq. (30), the result is[10]

$$\mu\int_{u_1}^{u}\frac{p'(v)dv}{(u-v)^{\frac{1}{2}}}\sim 1+\frac{2\Gamma(2/6)}{3\pi^{\frac{1}{2}}\Gamma(5/6)}\frac{b_1}{(u-u_0)^{1/6}}+\frac{\Gamma(1/6)}{3\pi^{\frac{1}{2}}\Gamma(4/6)}\frac{b_2}{(u-u_0)^{2/6}}+\left(\frac{2}{\pi}b_3-\mu\right)\frac{1}{(u-u_0)^{3/6}}-\frac{\Gamma(-1/6)}{3\pi^{\frac{1}{2}}\Gamma(2/6)}\frac{b_4}{(u-u_0)^{4/6}}-\frac{2\Gamma(-2/6)}{3\pi^{\frac{1}{2}}\Gamma(1/6)}\frac{b_5}{(u-u_0)^{5/6}}+2\frac{b_6}{(u-u_0)}. \tag{33}$$

[9] Higher terms are of the form $[\ln(u-u_0)]^m/(u-u_0)^{n/6}$, where m and n are integers.

[10] See the appendix for the evaluation of the integral.

By Eq. (30), this expression is also asymptotic to

$$1-\frac{1}{p^{\frac{1}{2}}}-\frac{1}{6p'}\frac{d}{du}(p^{7/3}p'^{2}). \tag{34}$$

If (34) is expanded by (32b), the coefficients of corresponding powers of $(u-u_0)$ may be equated to give a set of successive equations for the parameters $b_1, b_2, \cdots, b_6$. At each step, one has a linear equation for the unknown parameter. A tabulation of these parameters for various superheat conditions in water for an external pressure of 1 atmos is given in Table II..

The leading terms in the asymptotic solution are

$$p=(R/R_0)^3\sim\frac{2}{\pi\mu}u^{\frac{1}{2}}\{1+0(u^{-1/6})\},$$

$$t\sim\frac{3}{\alpha}\left(\frac{\pi\mu}{2}\right)^{4/3}u^{\frac{1}{2}}\{1+0(u^{-1/6})\},$$

$$T-T_0\sim-\frac{\alpha^2R_0^2}{A}\{1+0(u^{-1/6})\}.$$

Thus,

$$R\sim R_0\left(\frac{2}{\pi\mu}\right)\left(\frac{\alpha t}{3}\right)^{\frac{1}{2}}\{1+0(t^{-\frac{1}{2}})\}, \tag{35}$$

$$T-T_0\sim-\frac{\alpha^2R_0^2}{A}\{1+0(t^{-\frac{1}{2}})\}. \tag{36}$$

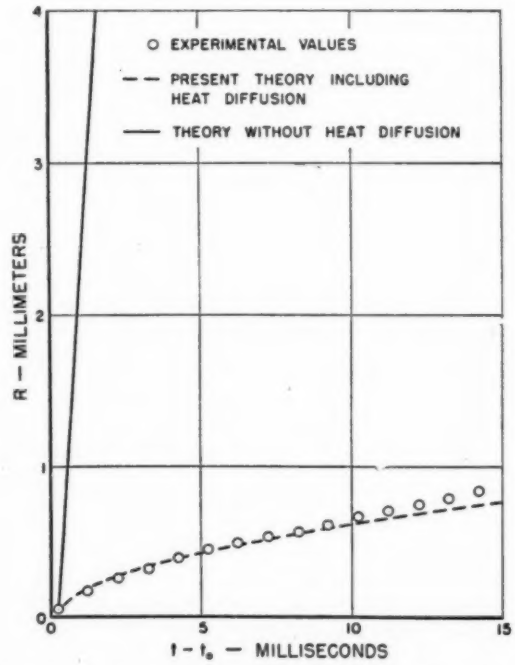

O EXPERIMENTAL VALUES
PRESENT THEORY INCLUDING HEAT DIFFUSION
THEORY WITHOUT HEAT DIFFUSION
R – MILLIMETERS
t – t₀ – MILLISECONDS

FIG. 2. Comparison of theoretical bubble radius-time values with experimental values in water at 1 atmos with a superheat temperature of 103.1°C. The solid curve is the Rayleigh growth curve neglecting the cooling effect; the dashed curve includes this effect by the method indicated in the text.

The vapor density ρ' may be approximated by a linear function of the temperature over the small range of present concern. Consequently,

$$\rho' \sim \text{const} \cdot \{1+0(t^{-\frac{1}{2}})\} \quad \text{as} \quad t \to \infty,$$

by Eq. (36). The volume of the bubble, on the other hand, behaves like

$$R^3 \sim \text{const} \cdot t^{\frac{1}{2}} \{1+0(t^{-\frac{1}{2}})\} \quad \text{as} \quad t \to \infty.$$

It is evident that the vapor density varies slowly in the asymptotic range compared with the bubble volume so that the neglect of $d\rho'/dt$ compared with dR^3/dt in the formulation of the boundary condition of Eq. (11) is justified.

III. COMPARISON WITH EXPERIMENT AND CONCLUSION

With the values of the constants given in Table II, radius-time curves have been computed for water at 1 atmos for superheats in the range of 102°C to 106°C. These curves are presented in Fig. 1. Inasmuch as the delay period may be chosen arbitrarily so far as these asymptotic solutions are concerned, the time scale is determined only within an arbitrary constant which varies from one curve to another. The actual spacing of the curves as presented was chosen so that the time intercepts at $R=0.04$ cm were equally spaced.

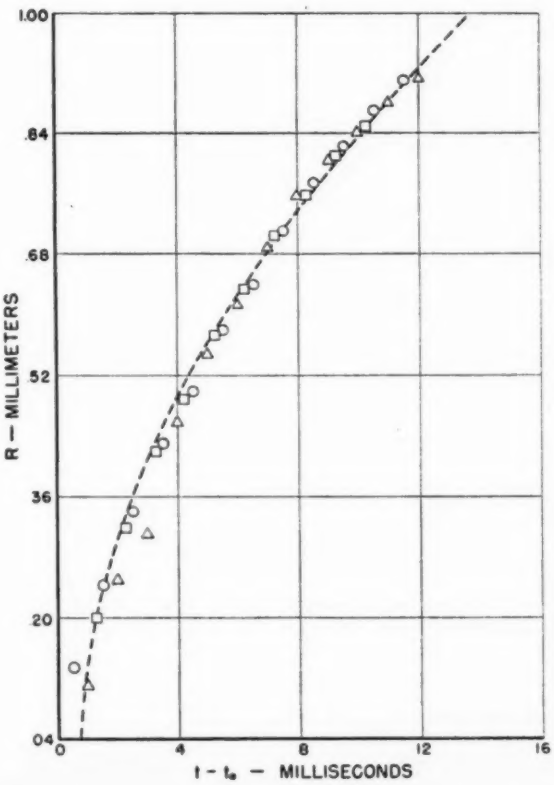


FIG. 3. Comparison of theoretical radius-time values with three sets of experimental values obtained in water superheated to 104.5°C at an external pressure of 1 atmos.

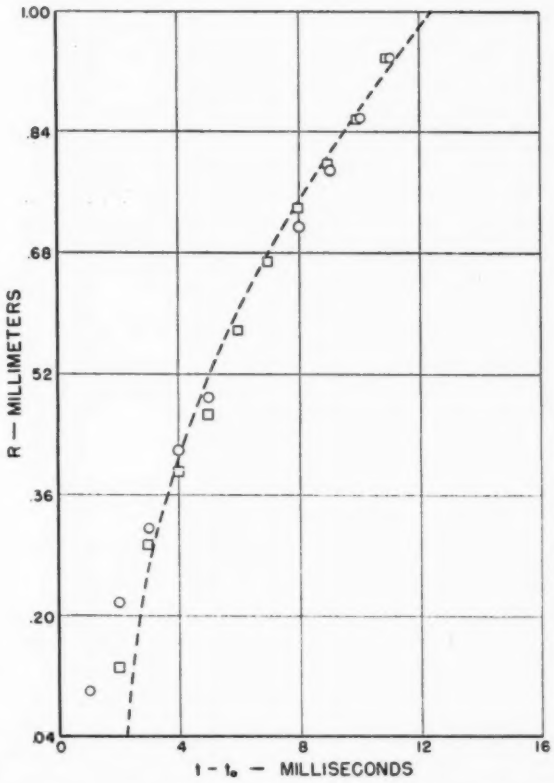


FIG. 4. Comparison of theoretical radius-time values with two sets of experimental values obtained in water superheated to 105.3°C at an external pressure of 1 atmos.

Observations have been made by Dergarabedian[11] on the growth of vapor bubbles in superheated water. The comparison between the theoretical curves and the observed values is shown in Figs. 2, 3, and 4. The theoretical curves were obtained by interpolation from the values graphed in Fig. 1. The time origins for both the theoretical curves and the experimental points are both arbitrary so that a time translation of the theoretical curve has been made in each case to give the best fit. The agreement is, however, seen to be very good. The importance of the cooling effect is evident from Fig. 2 where the Rayleigh solution [Eq. (7)] is also shown.

Equation (35) gives for the leading term in the asymptotic velocity

$$\dot{R} \sim \frac{R_0}{\pi\mu}\left(\frac{\alpha}{3t}\right)^{\frac{1}{2}} = \left(\frac{3}{\pi}\right)^{\frac{1}{2}} \frac{k(T_0-T_b)}{L\rho' D^{\frac{1}{2}}} \frac{1}{t^{\frac{1}{2}}}. \tag{37}$$

Although the number of terms given in Eq. (32b) is sufficient for high accuracy in the range of values of R used in Figs. 1–4, the leading term in Eq. (32b) is in error by approximately 10 to 30 percent for the smallest values of R. This leading term is sufficiently simple, however, that a rough physical argument should give the essential variation of Eq. (37). The argument is as

[11] P. Dergarabedian, J. Appl. Mech. (to be published).

follows: At a time t at which the bubble radius R is much greater than R_0, the difference between the temperature in the liquid at the bubble wall and that in the liquid at a distance is only slightly less than T_0-T_b. This temperature drop takes place principally in a liquid layer around the bubble of approximate thickness given by the diffusion length $(Dt)^{\frac{1}{2}}$. The heat flow into the bubble per unit time is therefore given roughly by

$$\dot{Q}\approx[k(T_0-T_b)/(Dt)^{\frac{1}{2}}]4\pi R^2. \tag{38}$$

The heat requirement per unit time of evaporation, on the other hand, is

$$\dot{Q}=Ld/dt(4\pi R^3\rho'/3)\approx 4\pi R^2\dot{R}L\rho'. \tag{39}$$

When (39) is equated to (38), there results

$$\dot{R}\sim k(T_0-T_b)/L\rho'(Dt)^{\frac{1}{2}}, \tag{40}$$

which agrees in order of magnitude with the leading term as given by Eq. (37).

Some experiments have recently been performed by Dergarabedian on vapor bubble growth in pure CCl_4. At moderate superheats, its vapor pressure curve is approximately parallel to that of water, displaced to lower temperatures. If the rates of bubble growth in these two liquids are compared at the same value of the temperature difference (T_0-T_b), they should thus be roughly in the same ratio as $k/(L\rho'D^{\frac{1}{2}})$ for the two liquids. This constant is about 3.5 greater in water than in carbon tetrachloride. Dergarabedian's observations on bubble growths are in good agreement with this value.

APPENDIX

Evaluation of the Temperature Integral [Eq. (33)]

By differentiating $p(u)$ [Eq. (32b)] and substituting into the temperature integral there results, after a change of variable,

$$\mu\int_{u_1}^{u}\frac{p'(v)dv}{(u-v)^{\frac{1}{2}}}$$

$$\sim\frac{1}{\pi}\int_{(u_1-u_0)/(u-u_0)}^{1}\frac{dv}{v^{\frac{1}{2}}(1-v)^{\frac{1}{2}}}\left\{1+\left[\frac{2}{3}\frac{b_1}{(u-u_0)^{1/6}}\right]\frac{1}{v^{1/6}}\right.$$

$$+\left[\frac{1}{3}\frac{b_2}{(u-u_0)^{2/6}}\right]\frac{1}{v^{2/6}}-\left[\frac{1}{3}\frac{b_4}{(u-u_0)^{4/6}}\right]\frac{1}{v^{4/6}}$$

$$-\left[\frac{2}{3}\frac{b_5}{(u-u_0)^{5/6}}\right]\frac{1}{v^{5/6}}$$

$$\left.-\left[\frac{b_6}{u-u_0}\right]\left[\ln v+\ln(u-u_0)-2\right]\frac{1}{v}\right\}. \tag{a}$$

Consider a typical integral appearing in (a),

$$I_s(x)=\int_x^1\frac{v^{s-1}dv}{(1-v)^{\frac{1}{2}}}. \tag{b}$$

For $0<x<1$, $\mathrm{Re}(s)>0$, this is the hypergeometric function[12]

$$I_s(x)=\frac{1}{s}[F(\tfrac{1}{2},s;s+1;1)-x^sF(\tfrac{1}{2},s;s+1;x)], \tag{c}$$

$$\sim\frac{\Gamma(\frac{1}{2})\Gamma(s)}{\Gamma(s+\frac{1}{2})}-\frac{1}{s}x^s \quad \text{as} \quad x\to 0^+. \tag{d}$$

By an argument based on the theory of analytic continuation, however, it may readily be shown that (c) is valid provided only that $0<x<1$, $s\neq 0, -1, -2, \cdots$, and hence that (d) holds for $0<x<1$, $\mathrm{Re}(s)>-1$. For $s=0$, (d) is meaningless, but gives a correct result under a limiting procedure. By differentiating (d) with respect to s at $s=-\frac{1}{2}$, one readily finds

$$\int_x^1\frac{\ln v\,dv}{v^{\frac{1}{2}}(1-v)^{\frac{1}{2}}}\sim -2\pi+\frac{2}{x^{\frac{1}{2}}}(\ln x+2) \quad \text{as} \quad x\to 0^+. \tag{e}$$

With the aid of (d), (e), Eq. (a) therefore becomes

$$\mu\int_{u_1}^{u}\frac{p'(v)dv}{(u-v)^{\frac{1}{2}}}$$

$$\sim\left\{1+0.89266\frac{b_1}{(u-u_0)^{1/6}}+0.77306\frac{b_2}{(u-u_0)^{2/6}}\right.$$

$$+0.47545\frac{b_4}{(u-u_0)^{4/6}}+0.27450\frac{b_5}{(u-u_0)^{5/6}}$$

$$\left.+2\frac{b_6}{(u-u_0)}\right\}-\frac{2}{\pi}\frac{1}{(u-u_0)^{\frac{1}{2}}}(u_1-u_0)^{3/6}$$

$$\times\left\{1+\frac{b_1}{(u_1-u_0)^{1/6}}+\frac{b_2}{(u_1-u_0)^{2/6}}+\frac{b_4}{(u_1-u_0)^{4/6}}\right.$$

$$\left.+\frac{b_5}{(u_1-u_0)^{5/6}}+b_6\frac{\ln(u_1-u_0)}{u_1-u_0}\right\} \tag{f}$$

after a rearrangement, several terms cancelling. The Γ functions appearing in (d) have been evaluated numerically in (f). By comparison with the asymptotic form for $p(u)$ of Eq. (30), the last term in (f) may be written

$$1/(u-u_0)^{\frac{1}{2}}\{(2/\pi)b_3-\mu p(u_1)\}.$$

Inasmuch as $p(u_1)=1$ by assumption, Eq. (f) reduces to Eq. (33) of the text.

[12] E. T. Copson, *Introduction to Theory of Functions of a Complex Variable* (Oxford University Press, London, 1935), pp. 249, 251.

Contact Charging Between a Borosilicate Glass and Nickel*

JOHN W. PETERSON
Department of Physics, University of California, Berkeley, California
(Received September 28, 1953)

The charging of borosilicate glass spheres rolling on clean nickel has been studied under controlled conditions of cleanliness, humidity, and gas pressure. Strong evidence indicates the process of charge transfer to be a type of contact electrification akin to that operating between two metals. The observed charging is, however, influenced by other factors. The amount of charge acquired increases with surface areas in contact and hence with distance rolled until other factors intervene. The rate of charging depends on rolling speed and surface conductivity of the borosilicate glass, while the *maximum equilibrium* charge depends on the pressure of the ambient gas. The gas pressure effect is shown to be caused by electrical discharge of the highly charged sections of the borosilicate glass to the metal, as suggested by E. W. B. Gill, and this causes a minimum equilibrium charge to occur about 1 mm air pressure, as observed by D. E. Debeau. The effect of transverse electrical fields reported by E. W. B. Gill and G. F. Alfrey is negligible for low surface conductivity but becomes important at higher surface conductivity, in particular above 40 percent relative humidity at 20°C.

THE phenomenon of electrical charging between different materials must be one of the oldest known to mankind. It also appears to be one of the most complex, and the story of its development is largely one of contradictions. Much confusion, however, is eliminated by recognizing that more than one type of charging may be active in a particular case. Thus, a certain order can be brought to the subject by grouping various types of charging according to the mechanism to which each is due.

L. B. Loeb[1] was the first to indicate this difficulty and grouped such charging into five classes: charging by electrolytic effects, contact charging between metals, spray electrification, the segregation of ions or electrons in gases, and tribo-electric or frictional charging† (that occurring where at least one of the materials is a dry insulator).

This paper is concerned only with the last class, tribo-electric charging. Much work has been done on this subject, most of which is contradictory and inconclusive, and excessive space would be required to give even a sketchy history of its development.‡ However, the gradual accumulation of evidence has led to the rather general acceptance of the contact theory[2–4] for most cases of charging,§ as held by A. Volta[5] and H. L. F. Helmholtz.[6] Where relative motion between the objects increases charging, it is considered to be due to an increase in the effective area of contact. Fairly definite indications substantiating this viewpoint came from the work of J. W. Hansen[7] and W. B. Kunkel[8] in this laboratory in their study of charging of dusts.

One of the first studies to give quantitative results which were reproducible and unambiguous, was that of D. E. Debeau[9] in 1924. Using sized quartz and rock-salt powder in a nickel system, he studied the effect of the atmosphere in which charging takes place. Careful control showed that charging depends upon both pressure and nature of the atmosphere. He established that not only are these variables important to the process of charging, but also that they must be controlled, if meaningful results are to be obtained.

In 1949 Gill and Alfrey[10] first reported that the charging of insulating particles in contact with a metal is strongly influenced by the application of an electric field perpendicular to the metal surface. Their results indicated that the effect was approximately linear in the applied field, and that a sufficiently strong electric field could reverse the sign of the charge acquired by the insulating material. In fact, ignoring modern theories of the solid state, Gill has ascribed all contact charging to the action of electrical fields without designating the nature of such fields.

This work and an earlier letter to *The Physical*

* This work was supported by funds from the U. S. Office of Naval Research.

[1] L. B. Loeb, Science, **102**, 573 (1945).

† Also variously called frictional electrification, static electrification, and contact charging, to list some of the names used in the past.

‡ A bibliography of some hundred or so titles dealing with this subject has been prepared by J. W. Hansen ("The electrostatic charge on insecticidal dusts," dissertation, University of California, Berkeley, 1948) but the results recorded lead to little clarification of the processes at work, so they will not be listed.

[2] E. Thomson, J. Am. Inst. Elec. Engrs. **41**, 342 (1922).

[3] H. F. Richards, Phys. Rev. **22**, 122 (1923).

[4] P. W. Burbridge, Trans. Proc. New Zealand Inst. **59**, 663 (1928).

§ Most such charging appears to require only physical contact and not friction, and the nature of the charging appears to be similar to the contact charging of metals. Therefore, it seems that where applicable contact charging (as used, for example, by Debeau (see reference 9) and H. F. Richards (see reference 3) would be a more appropriate term than frictional charging or tribo-electric charging. Such a designation, in addition to describing the subject more accurately, would not have the misleading implication which appears to be associated with the word friction. Certainly, it would be a great convenience if such a single designation were generally agreed upon, so that, for example, all abstracts of papers on the subject could be brought under a single heading.

[5] E. Perucca, Z. Physik **51**, 268 (1928).

[6] H. L. F. Helmholtz, Ann. Physik. **7**, 337 (1879).

[7] J. W. Hansen, Phys. Rev. **72**, 741 (1947); **73**, 532 (1948).

[8] W. B. Kunkel, J. Appl. Phys. **21**, 820 (1950).

[9] D. E. Debeau, Phys. Rev. **66**, 9 (1944).

[10] E. W. B. Gill and G. F. Alfrey, Nature **163**, 172 (1949).

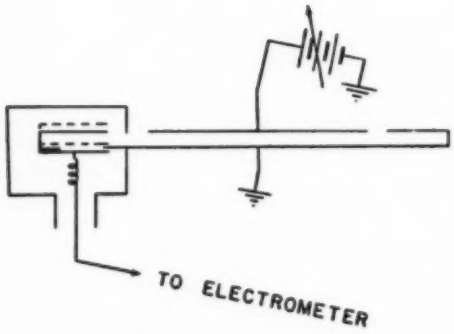


FIG. 1. Schematic diagram of apparatus.

Review by Gill[11] influenced the direction of the present investigation, and as it progressed, the knowledge accumulated seemed clearly to indicate the processes at work. It appeared to confirm the conclusion of previous workers‖ that such charging consists in the transfer of electrons between an insulator and metal in contact, and the conclusion of Loeb and Kunkel[5] that the charging of dusts is due to the transfer of electrons or ions across the surface of contact.

In the present work, in order to have smooth rolling and a small, well-defined area of contact, two-mm borosilicate glass spheres¶ were used, rolling on nickel in a carefully controlled atmosphere. The apparatus is shown schematically in Fig. 1. Nickel was chosen for its construction because of its slow rate of oxidation and favorable outgassing properties. The glass spheres rolled on the lower of a pair of concentric cylindrical nickel plates with a one-centimeter gap, which constituted a segment of a cylindrical condenser. The lower plate was grounded, and any desired potential up to ± 1200 volts was applied to the upper plate, producing a field of ± 1050 volts per centimeter at the surface of the lower plate. The two plates were separated by quartz spacers held in place by small bolts insulated with quartz tubing. The curved cross section of the lower plate served to confine the spheres to the approximately uniform field region. The spheres could be held in a cup at one end of the lower plate, and then the whole apparatus rotated about a horizontal axis, allowing the spheres to roll through the high field region to the other end. Here, the spheres were collected in a Faraday cage and the charge measured. The cup which constituted the Faraday cage could then be rotated by magnetic control until it rested on the upper surface of the lower plate. From this position the spheres were returned by rotating the entire apparatus again, allowing them to roll back to the holding cup. When the receiving cup had been rotated back to the receiving position, the apparatus was ready for another run. The receiving cup and measuring circuit were all carefully shielded in order to avoid the effects of stray fields. The entire apparatus was enclosed in a glass envelope with a ground-glass joint leading to the pumping system in order to allow the necessary rotation.

The pumping system consisted of a mercury diffusion pump, backed by a high-speed mechanical pump with a liquid-nitrogen-cooled trap between the experimental tube and the diffusion pump. This system produced vacuums of the order of 3×10^{-6} mm Hg in the experimental tube. Pressures from 1×10^{-5} to 3 mm Hg were measured by the use of a double McLeod gauge, and higher pressures by a manometer. The tube was filled to the desired pressure with air, which had been slowly passed through a 400-centimeter coil trap cooled by frozen ethyl alcohol. The temperature of the trap thus was high enough to prevent trapping oxygen out of the air while still effectively trapping water vapor and carbon dioxide, which are the principle contaminants to be avoided.

Initial measurements were made both in vacuum and dried air, and showed a much smaller effect of the electric field than those found by Gill and Alfrey. As might be expected, the charge of the sphere, which normally was negative, was increased when the lower plate was negative with respect to the upper, and decreased in the reverse case. The change in charge was found to be linear with the applied electric field, within experimental accuracy. Therefore, results are given in terms of Δq, the difference between the charge measured when the upper plate was at $+1200$ volts, and that measured with the upper plate at -1200 volts. Appreciable values of Δq were found only after water vapor had been intentionally admitted to the tube. After this observation, a systematic study was made of the variation of Δq with relative humidity, the results of which are shown in Fig. 2.

Relative humidity was varied by evacuating the tube to about 10^{-6} mm Hg, and then connecting it to a reservoir of distilled water until equilibrium had been reached. The reservoir was held at such a temperature as to give the desired vapor pressure of water in the tube. After measurements had been made, dry air was

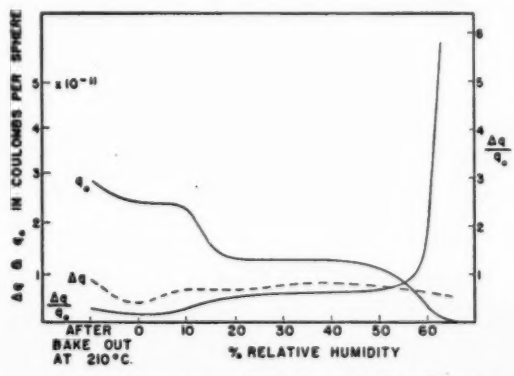


FIG. 2. Δq and q_0 as functions of relative humidity.

[11] E. W. B. Gill, Phys. Rev. **74**, 842 (1948).

‖ No attempt has been made to obtain a complete list of references for this hypothesis. References 3, 4, and 5 are representative.

¶ Smaller spheres tended to stick to the metal. In light of the work of F. P. Bowden and J. E. Young [Proc. Roy. Soc. (London) **A208**, 311 (1951)] this sticking can be attributed to "welding" of areas of the spheres to the metal under the great pressure produced by image forces between the charged sphere and the metal.

admitted, producing a range of pressures up to atmospheric. The pressure of air within the tube had no significant effect upon the charging so results shown in Fig. 2 are in terms of relative humidity only. The total charge in zero field (designated q_0) is also shown. In addition, to emphasize the effect of water vapor the $\frac{\Delta q}{q_0}$ quantity—is shown. The point to the left of zero relative humidity represents measurement taken after outgassing at 210°C.

The variation of Δq and q_0 with relative humidity appears to be due to variations in the amount of water adsorbed upon the surface of the sphere, with the accompanying variation in surface conductivity.**

Borosilicate glass is known to adsorb water heavily, producing quite thick layers at high relative humidity. This adsorbed water is not all removed by evacuation but requires heating to above 300°C. to be completely driven off.[12] Consequently, surface conductivity increases with relative humidity, as the adsorbed water layer becomes thicker. Outgassing appeared to drive off some of the strongly adsorbed water molecules, and decreased the surface conductivity.

The decrease in q_0 with increasing surface conductivity can be explained by considering the potential distribution over the surface of the sphere. While the charge on the sphere can be considered to be rather uniformly distributed over its surface after prolonged rolling, the potential distribution will not be uniform. This is evident by analogy with a parallel plate condenser, where potential difference increases with increasing separation of the plates, if the charge density remains constant. Thus, that portion of the sphere which is close to the metal will have a low potential, whereas areas which are relatively far from the metal surface will have a relatively high potential. There is, therefore, a strong potential gradient, tending to drive charge downward, toward the point of contact. It, therefore, follows that increasing surface conductivity will facilitate this flow of charge, with the result that less charge remains on the surface of the spheres.

The applied electric field increases this normal charge flow when the upper plate is negative and opposes it in the opposite case (since the charge on the sphere is negative). Thus, Δq is a measure of the effectiveness of the applied field in modifying the normal charge flow and would be expected to increase with conductivity.

Since q_0 decreases with increasing conductivity, so do the potential gradients to which it gives rise. Therefore, a point is reached at which the applied field becomes greater than the fields due to the charges on the spheres, and Δq becomes approximately equal to q_0.

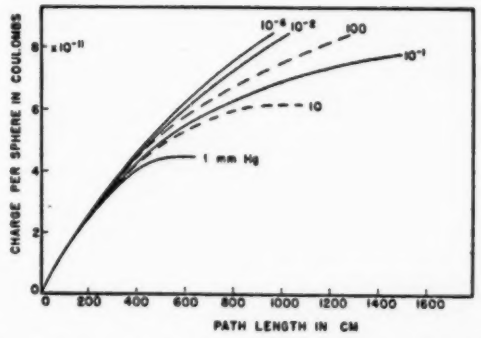


FIG. 3. Charge acquired as a function of path length for various pressures of air.

As conductivity increases further, Δq increases and q_0 decreases, until the spheres have become essentially conducting and measurements have become meaningless. It is probable that at very high conductivity, much of the charge was lost from the spheres before they had passed between the high field region and the measuring cup. Therefore, the measured values of Δq are probably too low for high values of relative humidity.

Δq had small but appreciable values even in the case of lowest surface conductivity. This is attributed to the polarization surface charge†† present at the surface of a dielectric in an electric field. The measured value of Δq for this case corresponds approximately to one fourth the maximum polarization surface charge density to be expected on a glass sphere in a uniform field. This makes the explanation seem reasonable in view of the uncertainty in the actual value of P_n and of the amount of surface charge which is retained by a sphere. The minimum in Δq is considered to occur when the polarization charge begins to leak away, before the surface has become sufficiently conducting for the applied field to have an appreciable effect upon the charge distribution.

Since it had been found that the measured charge varied with the speed of rolling, these measurements were all made at the lowest possible speed. The spheres were rolled back and forth along the lower plate at least ten times, which was sufficient to give a saturation charge uniformly distributed over the surface of each sphere. Careful control of the speed of rolling led to reproducibility within a probable variation of not more than five percent.

The increase of charge with speed of rolling is also considered to be a result of the potential distribution over the sphere. Since the potential of an area of the sphere increases with its separation from the metal surface, there is a field tending to drive newly acquired charge back toward the point of contact. Therefore, the faster the sphere rolls, the less time there is available for the resulting current to drain charge from the element of area. Thus, at higher speeds an increasing proportion of charge remains on the sphere. While

** This explanation is at variance with the proposal of W. R. Harper [Nature **167**, 400 (1951)] that gaseous discharge was responsible for the results observed by Gill and Alfrey.

[12] S. Dushman, *Vacuum Technique* (John Wiley and Sons, Inc., New York), Chap. 6, Sec. 8.

†† Surface charge density $= P_n$, the normal component of polarization.

speed of rolling was poorly controlled in this experiment, there was no doubt that the effect of speed was very real.

Further interesting results were obtained by plotting charge acquired by the spheres when rolling at a high speed as a function of the distance rolled. Figure 3 shows such curves for various pressures of dry air. Within experimental accuracy these curves all had the same initial slope, although the pressure varied from 10^{-5} mm Hg to 700 mm Hg. This was considered significant because Debeau's work had indicated a variation of the charging process with pressure of the atmosphere.‡‡ If there were such a variation, it would be expected to result in a variation of the rate of acquiring charge which was not observed. There was, however, a strong pressure dependence of the saturation charge which appears to agree qualitatively with Debeau's results. It seems evident, therefore, that Debeau was, in fact, measuring saturation charge as a function of pressure.

In 1949 Gill[11] proposed an alternative explanation of Debeau's results; namely, that gaseous discharge from the insulating particles to the metal surface limited the amount of charge which could be retained on the insulators. Certainly, the pressure dependence of charge, as found by Debeau, does have a striking resemblance to the pressure dependence of sparking potential for a pair of electrodes. This explanation is compatible with the observed fact that the initial rate of charging is independent of pressure, since gaseous discharge can begin only after the sphere has acquired sufficient charge to produce the required potential. Further evidence is supplied by the large fluctuations in charge observed after the maximum had first been reached. Such fluctuations can easily be explained by assuming that a gaseous discharge once initiated drains a considerable charge from the spheres. Therefore, the measured charge is reduced, and then gradually increased again in the course of rolling. Thus, all pertinent evidence from this experiment points to gaseous discharge from the insulating particles as being responsible for the pressure variation of charging.

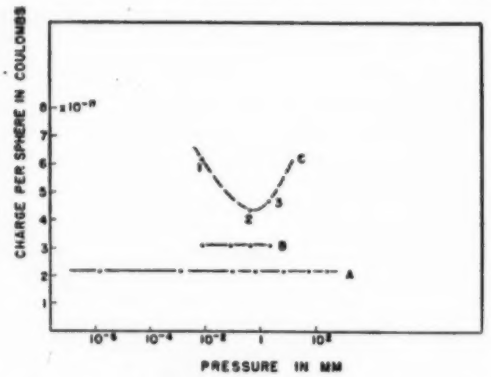


FIG. 4. Maximum charge as a function of pressure of air. (*A*) Low speed, before outgassing. (*B*) Low speed, after outgassing. (*C*) High speed, after outgassing.

‡‡ Debeau (see reference 9) interpreted his results in terms of changes in surface properties which resulted from variations with pressure of the adsorbed gas layers.

Figure 4 provides an interesting example of the effects of surface conductivity and gaseous discharge in limiting the total charge of the spheres. Curves *A* and *B* represent measurements made in vacuum at low rolling speed before and after outgassing at 210°C. At such speed surface conductivity limits charge to values below that required for gaseous discharge, and maximum charge is independent of pressure. The measured charge after outgassing was higher because of the lower conductivity resulting from cleaner surfaces, and, therefore, Curve *B* lies above Curve *A*.

During the course of taking the data for Curve *B*, Points 1, 2, and 3 of Curve *C* were obtained by high-speed rolling. As discussed previously, high rolling speed results in an increased charge on the sphere, which in this case was sufficient to produce gaseous discharge. Therefore, at such speeds the maximum charge is determined by gaseous discharge.

The charge transfer which gives rise to contact charging is considered to result from a difference in the effective work functions of the insulator and the metal. Thus, the effect is considered to be analogous to the contact potential difference between dissimilar metals.§§ Since the borosilicate glass charged negatively in contact with nickel, its effective work function must be greater than that of nickel. Therefore, an increase in work function of the nickel would be expected to decrease charging by decreasing the work function difference. Such a decrease was observed after the work function of the nickel was increased by oxidation in a glow discharge and, therefore, offers support for this hypothesis.

In summary, this experiment accomplished two purposes. First, the effect of a transverse applied electric field upon contact charging was shown to have a logical explanation and to be important only for high surface conductivity. Secondly, new light was thrown upon the general subject of contact charging. Conduction over the surface was shown to be an important factor in limiting maximum charge, as evidenced by variation of charging with speed of rolling. The process of acquiring charge was shown to be independent of pressure, and strong evidence was given for the proposal that pressure variation of maximum charge is due to gaseous discharge. Further, some evidence was obtained in support of the work function difference hypothesis for the transfer of charge from a metal to an insulator. Finally, this work showed that surface conditions, gas pressure, and speed of rolling must all be controlled in order to obtain reproducible results.

The author wishes to express his sincere thanks, and to acknowledge his very great debt to Professor Leonard B. Loeb, who suggested this problem and under whose guidance its study was carried out.

§§ W. R. Harper [Proc. Roy. Soc. (London) **A205**, 83 (1951)] has shown that the charges acquired by metallic spheres in contact can be accurately predicted from their measured contact potential differences.

Estimating Differential Scattering Cross Sections in Gases

TINO AHRENS
Nuclear Physics Group, Consolidated Vultee Aircraft Corporation, Fort Worth, Texas
(Received September 28, 1953)*

A simple method of estimating differential angular scattering cross sections in an infinite medium is presented. It assumes that single scattering theory approximates the radiation flux at a detector which is separated from the source by much less than a mean free path.

INTRODUCTION

THE once scattered differential flux $f(D;r,\theta)drd\theta$ at a point detector due to a monoenergetic isotropic unit point source is expressed by

$$f(D;r,\theta)drd\theta=(1/4\pi r^2)n\delta(\alpha)2\pi r^2\sin\theta drd\theta/\rho^2 \quad (1a)$$

if attenuation is neglected. The symbols are defined as follows: D=source detector separation; n=number of scattering centers per cm^3; θ=polar angle of emission with respect to the source detector axis; α=scattering angle; r=distance from source to differential scattering volume $dV=2\pi r^2\sin\theta drd\theta$; ρ=distance from differential scattering volume dV to detector; $\delta(\alpha)$=differential angular scattering cross section or this cross section times the energy of the radiation after scattering if the energy flux is measured rather than the (number) flux. Integration of (1a) over all space may in general constitute a good approximation to the answer of the problem of obtaining the flux at a distance from the source of much less than a mean free path after the radiation has undergone scattering by an essentially infinite medium.[1]

To accomplish this integration it is convenient, and conventionally done, to transform the integrand from r dependence to α dependence. Thus, the transformation $r=D(\cos\theta-\mathrm{ctg}\alpha\sin\theta)$ and the relation $1/\rho=\sin\alpha/D\sin\theta$ must be called upon. The flux can now be represented by the integral.[2]

$$\mathfrak{F}(D)=\int_0^\pi F(D;\theta)d\theta, \quad (2a)$$

where

$$F(D;\theta)=(n/2D)\int_\theta^\pi\delta(\alpha)d\alpha. \quad (3a)$$

THEORY

Differentiation of (3a) with respect to θ renders

$$dF(D;\theta)/d\theta=-n\delta(\theta)/2D. \quad (4a)$$

This result suggests an experimental setup where the source is provided with a shield permitting radiation to emanate in form of a cone with variable opening angle 2θ. This can, of course, be accomplished, e.g., through building the, say spherical, shield up of removable conical shells of angular thickness $\Delta\theta$. A "point" detector near by, protected from the direct beam, will then measure successively the scattered fluxes $\int_0^{\Delta\theta}F(\theta')d\theta'$, $\int_0^{2\Delta\theta}F(\theta')d\theta'$, etc. The forthcoming data are plotted in the form of $F(\theta)$ *versus* θ. The slope of the resulting curve is a measure of the angular differential scattering cross section $\delta(\theta)$.

Taking account of attenuation within the framework of single scattering lets the differential flux be expressed by

$$f(D;r,\theta)=ne^{-\mu r}\delta(\alpha)\sin\theta drd\theta/2\rho^2, \quad (1b)$$

μ being the total macroscopic interaction coefficient comprising both absorption and scattering. Inclusion of the attenuation $e^{-\mu(\alpha)\rho}$ which the radiation suffers between the point of first scattering and the detector would, at least for energies where the importance of scattering overwhelms that of absorption, imply a partial consideration of second scattering; however second scattering, the total effect of which is of a smaller order of magnitude under the present circumstances, shall be neglected in this discussion. Our function $F(D;\theta)$ will therefore be described by

$$F(D;\theta)=(n/2D)\int_\theta^\pi e^{-\mu D(\cos\theta-\mathrm{ctg}\alpha\sin\theta)}\delta(\alpha)d\alpha. \quad (3b)$$

Differentiating as before with respect to θ furnishes us with the relation

$$\delta(\theta)=\int_\theta^\pi[d(e^{-\mu D(\cos\theta-\mathrm{ctg}\alpha\sin\theta)})/d\theta]\delta(\alpha)d\alpha - (2D/n)dF(D;\theta)/d\theta, \quad (4b)$$

thus permitting to improve the initial values of $\delta(\theta)$ by iteration. Obviously θ, the polar angle of emission, may be replaced by θ_r, that of reception. One then faces

* The original unrevised manuscript was submitted July 18, 1953.

[1] Bernard, Miller, Jones, and Kunkel, "Air scattering of gamma radiation from cobalt 60" (to be published). The work of these authors demonstrates good agreement between predicted results of single scattering theory and experimental data as far as gamma rays of an initial energy of 1.25 Mev are concerned. In the case of neutrons, for whose benefit this paper is mainly written, the influence of multiple scattering may be somewhat larger but is still hoped to be small enough to allow application of the present treatment.

[2] As long as no attenuation terms are included the double integral (2a) can immediately be reduced to a single integral by changing the order of integration. One obtains $\mathfrak{F}(D)=(n/2D)\times\int_0^\pi\alpha\delta(\alpha)d\alpha$.

the expressions

$$\mathfrak{F}(D)=\int_0^{\pi} F_r(D;\theta_r)d\theta_r \qquad (2c)$$

and

$$F_r(D;\theta_r)=(n/2D)\int_{\theta_r}^{\pi} e^{-\mu D\sin\theta_r/\sin\alpha}\delta(\alpha)d\alpha \qquad (3c)$$

rather than (2a) and (3b).

Consequently (4b) goes over into

$$\delta(\theta_r)=e^{\mu D}\left\{\int_{\theta_r}^{\pi}\left[d(e^{-\mu D\sin\theta_r/\sin\alpha})/d\theta\right]\delta(\alpha)d\alpha - (2D/r)dF_r(D;\theta_r)/d\theta_r\right\}. \qquad (4c)$$

In the corresponding experimental setup the detector becomes directed instead of the source.

It seems particularly worth mentioning that, if willing to abandon the attenuation refinement, generalization to any point source of angular spectrum $A(\theta)$ and energy spectrum $E(\epsilon)$ is possible as long as angular and energy variables are separable. Then we have

$$F(D;\theta)=(nA(\theta)/2D)\int_{\theta}^{\pi}\langle\delta(\alpha)\rangle_{Av}d\alpha \qquad (3d)$$

with

$$\langle\delta(\alpha)\rangle_{Av}=\int_0^{\infty}\delta(\alpha,\epsilon)E(\epsilon)d\epsilon, \qquad (5)$$

$\delta(\alpha,\epsilon)$ being the cross section for radiation of energy ϵ. In this more general case $F(\theta)/A(\theta)$ is to be plotted *versus* θ; $\langle\delta(\alpha)\rangle_{Av}$ is a cross section peculiar to the energy spectrum of the particular source under consideration.

It gives me pleasure to record stimulating conversations with Mr. D. Selengut.

JOURNAL OF APPLIED PHYSICS VOLUME 25, NUMBER 4 APRIL, 1954

The Backscattering of the Co^{60} Gamma Rays from Infinite Media*

EVANS HAYWARD AND JOHN H. HUBBELL
National Bureau of Standards, Washington, D. C.

(Received September 8, 1953)

The radiation backscattered by semi-infinite slabs of wood and steel wool, when irradiated by the Co^{60} gamma rays, has been investigated by means of a sodium iodide scintillation spectrometer. The main contributions to the scattered radiation are produced in one or two Compton collisions. However, the radiation that has suffered so many collisions as to have an isotropic angular distribution has also been observed.

INTRODUCTION

THE purpose of this paper is to describe a series of experiments performed to investigate the properties of the radiation that emerges backward from a "semi-infinite" medium bombarded by the Co^{60} gamma rays. These gamma rays may be expected to undergo several Compton scatterings in an infinite medium before eventual photoelectric absorption. The photon is degraded in energy by each collision and this process, if it could continue indefinitely, would result in the concentration of the photons at lower and lower energies. However, the photoelectric effect becomes more important than Compton scattering at low energies, the transition energy depending on the atomic number of the material. As a result the lower energy photons are absorbed and the degradation by repeated Compton scattering is cut off.

The whole problem of the backscattering from an infinite medium is difficult to attack experimentally, and thus has been investigated very little. We have made a rather extensive series of measurements all of which are described in National Bureau of Standards Report No. 2264. Some similar work has recently been done by G. J. Hine[1] using a different geometry. A few of the results of our measurements are of general interest and are amenable to interpretation; examples of these data will be presented here.

EXPERIMENTAL DETAILS

We have used a single crystal NaI(Tl) scintillation spectrometer to look at the backscattered radiations, and though this instrument has some limitations for measuring continuous spectra quantitatively, it is certainly capable of pointing out the gross characteristics of the phenomenon. The response of sodium iodide to gamma rays of various energies has been discussed thoroughly in the literature,[2] but it seems worth while to mention here the specific limitations that it has for the problem at hand. The Co^{60} radiations consist of gamma rays at 1.17 and 1.33 Mev so that one would expect to observe photons of all lower energies. These scattered photons interact with the

* This work was supported by the U. S. Atomic Energy Commission and the U. S. Office of Naval Research.

[1] G. J. Hine, private communication.
[2] R. A. Hofstadter and F. A. McIntyre, Phys. Rev. 80, 631 (1950).

NaI crystal by photoelectric absorption and Compton scattering in varying amounts depending on their energy. Whenever photoelectric absorption takes place, the height of the pulse from the scintillation spectrometer is proportional to the energy of the incident photon (except just above the K edge of iodine at 33 kev). On the other hand, when Compton scattering occurs in the crystal, the pulses produced have a whole distribution of sizes, the maximum in this distribution corresponding to an energy slightly less than that of the incident photon. As long as the spectrum consists primarily of photons with energies between 40 and 150 kev, photoelectric absorption in the crystal predominates, and one can expect a one-to-one correspondence between photon energy and the pulse height that is measured. If, on the other hand, higher energy photons are present in appreciable numbers, then the measured pulse height distribution will not represent the actual spectrum, because some of the higher energy photons will undergo Compton scattering in the crystal and will appear degraded in energy. This type of spectrometer is therefore useful in measuring a continuous spectrum[3] below 150 kev in the absence of higher energies, and can be used to determine the presence of higher energy components.

The source was a ten-curie slug of Co^{60}. It was housed in a lead container, and the radiation was confined to a small angle by a lead collimator eight inches long having a three-quarter inch hole† (Fig. 1). The target was supported by a heavy steel frame capable of rotation so that the angle of incidence α could be changed, and the counter was arranged so that it could observe photons emerging from various parts of the target from various angles ϑ with respect to the incident beam.

Initially a pile of boards was used as the semi-infinite medium. Wood was chosen as the material for the scatterer not only because of its availability, but because of its low density which permits the scale of the phenomena to be spread out over a rather large region. A low atomic number material such as wood has the additional property of being nearly a pure Compton scatterer, since photoelectric absorption is not important except at extremely low energies. After some progress had been made in the experiment, it seemed desirable to construct a higher atomic number target having a density comparable to that of the wood pile. Photoelectric absorption would be more important in such a target, and the differences in the measurements could be directly attributed to this property. Such a target was made by compressing some large steel wool reels, so that their density was the same as that of the lumber. As far as the photons are concerned, steel wool represents a homogeneous medium because the strands of which it is composed are at least an order of magnitude smaller than the mean free path for Thompson scattering in iron. The target thickness of twenty in. was chosen to be in excess of that required to produce the maximum counting rate in the backward direction.

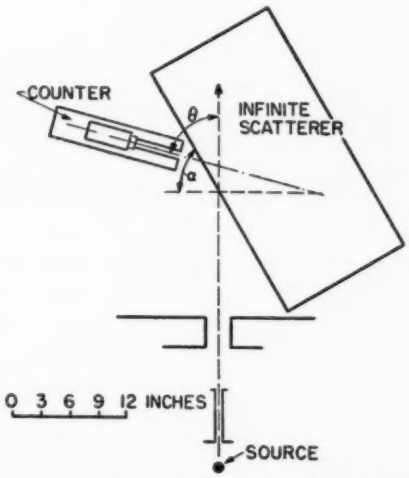


FIG. 1. A schematic diagram of the experiment.

The counter consisted of a NaI crystal approximately one-half in. in thickness and 1.5 in. sq with aluminum foil reflector, separated from a 5819 photomultiplier tube by a thin Lucite adapter. The tube, crystal, and preamplifier were housed in a copper cylinder. The top end was a collimator six in. long with a one-in. diameter hole, through which the crystal looked at a small portion of the target surface (see Fig. 1). A lead collimator was unsuitable because of the interference of its 75-kev fluorescence with the spectra under study, although an outside wrapping of lead around the copper was useful in reducing background. The preamplifier was a simple cathode follower from which the pulses were fed into a linear amplifier. Then they were analyzed by a sliding channel pulse-height analyzer.

The actual photon energies that were measured were in the range from 40 to 700 kev, so it was necessary to have on hand a number of sources having gamma rays in this region to be used as energy standards. These were Cs^{137} (662 kev), Hg^{203} (279 kev), with Tl x-rays at 73 kev, Cd^{109} (88 kev), and Na^{22}, which was used as a source of annihilation radiation. A calibration of the discriminator bias was made before, and usually after, the measurement of each pulse-height distribution using the appropriate source, or sources. The width of the calibration line also gives the resolution of the instrument at that energy. Under the actual running conditions of the experiment the resolution varied from about 15 percent at 662 kev to about 35 percent at 77 kev. This trend is in agreement with the work of Schardt and Bernstein.[4]

[3] L. Madansky and F. Rasetti, Phys. Rev. **83**, 187 (1951); T. B. Novey, Phys. Rev. **86**, 619 (1952); S. A. E. Johannesson, Arkiv Fysik **3**, 533 (1952).

† It was found necessary to use a one-eighth in. lead filter over the source to suppress a 200-kev component that originated when the primary gamma rays backscattered from the bottom of the container and the floor beneath.

[4] A. W. Schardt and W. Bernstein, Rev. Sci. Instr. **22**, 1020 (1951).

RESULTS

We have measured a series of pulse-height distributions using a geometry designed to show up any variations in the photon energy distributions as a function of angle in the backward direction. In one set of measurements the target was tipped so that the angle of incidence α was 60°. The counter was directed at a point ten cm along the surface from the point where the beam center entered the scattering medium, and adjusted to observe photons deflected through a series of angles in the interval from $\theta=50-150°$. The axes of counter and source were in the same plane.

Figure 2 shows an example of one of the distributions, (solid circles), for $\theta=70°$ taken using the wood target. It has been selected for purposes of discussion, and the following remarks apply as well to the distributions measured at the other angles as well as to those obtained using the steel wool target. The most outstanding features of this distribution are the maxima at 475 and 125 kev. The higher energy one is the photoelectric peak produced in the NaI crystal by the photons that have undergone one Compton collision in the target. The corresponding maxima were observed within two percent of the energy predicted by the Compton scattering law at all angles. The lower energy end of the pulse-height distribution is confused by the large number of pulses produced by Compton scattering of the harder photons in the crystal. In order to identify uniquely this part of the distribution, we have measured the pulse-height distributions, using a thin scatterer large enough in area to reproduce the geometry of the infinite scatterer. A thin target is one in which the photons undergo only single collisions, so that the measured pulse-height distribution is the one produced by the "monoenergetic" photons scattered within the angular interval defined by the detector. This spectrum consists of a photoelectric peak and Compton distribution. Figure 2 shows the pulse-height distributions corresponding to the thin (open circles), and thick (solid circles), targets normalized so that the areas under the photoelectric peaks are equal. The difference between the two distributions is the contribution from photons that have been scattered more than once in the target and consists of a continuous distribution of pulses with a broad peak at 120 kev. This whole distribution represents low enough energies so as to be very little affected by Compton scattering in the crystal.

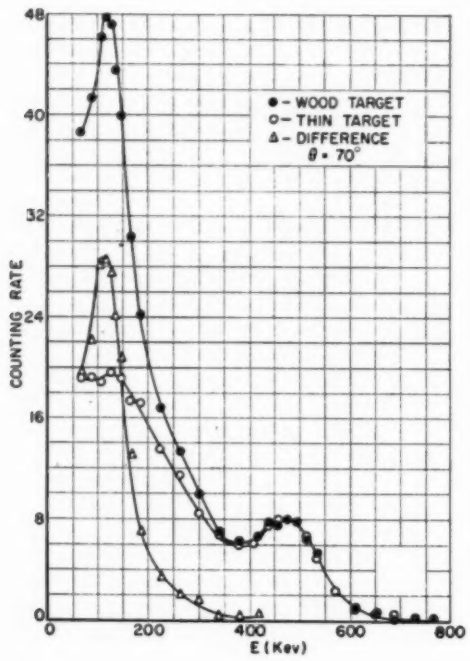


FIG. 2. The pulse-height distributions measured for a scatter angle of 70°. The solid circles represent the distribution obtained using the infinite wood scatterer, and the open circles obtained using a thin scatterer. These have been normalized so that the areas under the photoelectric peaks are equal. The third distribution (triangles) is the difference between the first two, and represents those photons that have suffered more than one collision.

This secondary maximum has four important properties that have been determined experimentally. (1) It does not come from single scattering. (2) The position of the peak goes up in energy as the scattering angle gets larger. (3) The peak is sharper for the small scatter angles, and becomes quite broad for scatter angles larger than 90°. (4) It disappears when the axes of the counter and source do not lie in a plane. An investigation of the simple processes able to produce such a maximum, led to the conclusion that it was to be expected on the basis of double scattering.

If a photon is to come from the source and be detected by the counter after two collisions, both scatters must take place in the plane occupied by the two collimators. With this simplification it can be shown that a large fraction of the pairs of large angle collisions having the same total deflection yield very nearly the same final photon energy. The observed peak is thus produced in the spectrum; it corresponds to the discontinuities in the energy distributions for the second scattered beam described in the paper by Spencer and Jenkins.[5] The pairs of angles that do not contribute to this peak simply produce a broad smear of energies that, along with the photons produced in higher orders of scattering, form the broad distribution over which the single and double scattering peaks rise. The energies, E_{ds}, at which these peaks should occur as a function of θ, were determined in a rough calculation and are given in Table I along with the energy, E_{ss}, of the singly scattered photon. The energy of the primary photon has been taken to be 1.25 Mev.

The way in which the width of this maximum varies with angle can also be predicted on the basis of two coplanar Compton scatters. In the extreme case of two 180° deflections, the photon would arrive with the

[5] L. V. Spencer and Fannie Jenkins, Phys. Rev. **76**, 1855 (1949).

minimum energy of 120 kev and a sharp line should be observed. As we decrease the magnitude of either or both scattering angles, the number of ways in which the photon can achieve its final direction after two scatters increases, and therefore the spread in energy of the twice scattered photons goes up. The positions of the maxima, as well as their variation in width agree very well with the experimental observations.

The pulse-height distributions measured, using both the wood and the steel wool targets, show the same qualitative features. The actual scattered intensities are even the same for photon energies above about 200 kev. Below this energy many of the photons are removed by the photoelectric effect in the steel wool target before they can escape.

Figure 3 shows the pulse-height distributions observed by the counter looking into the side of the target

TABLE I.

θ	E_{ss} (kev)	E_{ds} (kev)
50	667	122
70	479	122
90	362	132
120	268	155
150	225	190

in such a way that the counter could not see the axis of the beam. With this geometry the maxima discussed above are absent, and the photon distribution is that of the diffuse radiation, the radiation that has been scattered so many times that it has "forgotten" its original energy and direction. This broad distribution is isotropic and has been obtained by tipping the counter at several different angles relative to the target. The shape of the high-energy tail depends somewhat on the counter's orientation, since these photons arise in the lower orders of scattering. The differences in the spectra emitted by the wood and steel wool targets are striking and can be attributed to the photoelectric effect; both spectra have broad maxima at the energy when the buildup of low-energy photons produced in Compton scattering is counterbalanced by photoelectric absorption.

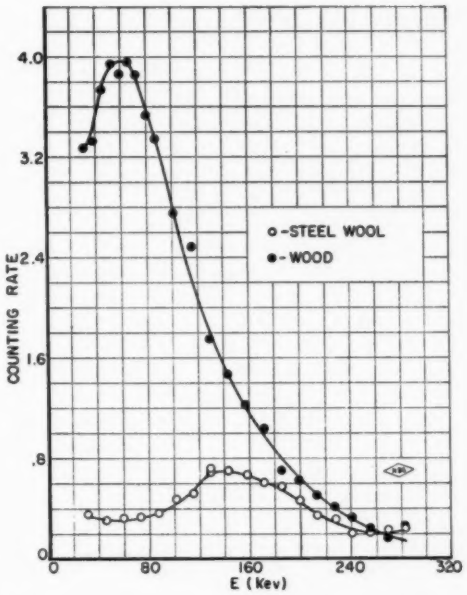


FIG. 3. These data represent the photon spectra obtained using the wood (solid circles), and the steel wool (open circles), targets with the counter adjusted so that its axis does not intersect that of the source radiation. These spectra result from photons that have undergone many collisions; the marked dissimilarity between the two distributions is a result of the difference in the photoelectric absorption for the two materials since the two infinite targets had the same density.

This type of investigation can be used to give semi-quantitative information about the energy spectra of backscattered photons at various angles, and positions of a semi-infinite medium. Because of the large number of integrations required and the rather large uncertainties in the solid angles subtended by the collimators, these results cannot be used to determine the reflection coefficient, which is the fraction of the radiation backscattered. For this reason we have not pursued the measurement any further; instead we have performed a "theoretical experiment"[6] using the Monte Carlo technique to determine the reflection coefficient of various material for 1-Mev photons.

[6] Evans Hayward and John H. Hubbell, Natl. Bur. Standards Report No. 2768.

Interpretation of Photoelastic Transmission Patterns for a Three-Dimensional Model*

D. C. DRUCKER† AND W. B. WOODWARD‡
Brown University, Providence, Rhode Island
(Received September 26, 1953)

The work of O'Rourke and Saenz on the photoelastic determination of quenching stresses is re-examined and extended. A brief general discussion is given on the determination of shear stresses or stress differences in the interior of a photoelastic model from retardation patterns obtained by passing light through the entire model. A method is proposed for the cylindrical case, with no variation in stress along the axis, which is feasible at present, while the proposal for the general axi-symmetric problem is feasible in principle. The fundamental difficulty of the general three-dimensional model is brought out but no solution is advanced.

INTRODUCTION

EACH of the standard procedures for three-dimensional photoelastic analysis has serious disadvantages. In a sense, both the scattering and the freezing techniques require slicing of the model. The physical slicing in the freezing method destroys the model and a new one is needed for each condition of loading. The alternative optical slicing, or probing, employing scattered light permits reuse of the model but the intensity of the scattered light is very low in most cases; exposure times are long and the accuracy of measurement is not always sufficient. Dynamic experiments, other than steady state problems, cannot be performed with either technique.

Clearly, the most desirable procedure would be one utilizing the retardation patterns obtained by passing polarized light through the entire three-dimensional model. Transient dynamic problems could be studied and the analysis of static problems made much more convenient. O'Rourke and Saenz[1] derived such a procedure for thermoelastic quenching stresses in long circular cylinders and spheres.[2] Although they assumed a very restrictive sum rule for the cylinder, $\sigma_z=\sigma_r+\sigma_\theta$ in the usual notation, their pioneer paper opened the possibility of great progress in photoelasticity.

The following discussion will remove the unnecessary restriction, extend the work of O'Rourke and Saenz, and discuss the over-all problem in greater generality. One objective is to obtain the maximum amount of photoelastic information with the use of as little "outside help" as possible. Viscous or plastic flow may take place in a model so that it is best not to depend upon the equations of elasticity. The equations of equilibrium may, on the other hand, be expected to hold but the later they are applied the better because the translation of optical effect into stress may also be in error due to flow.

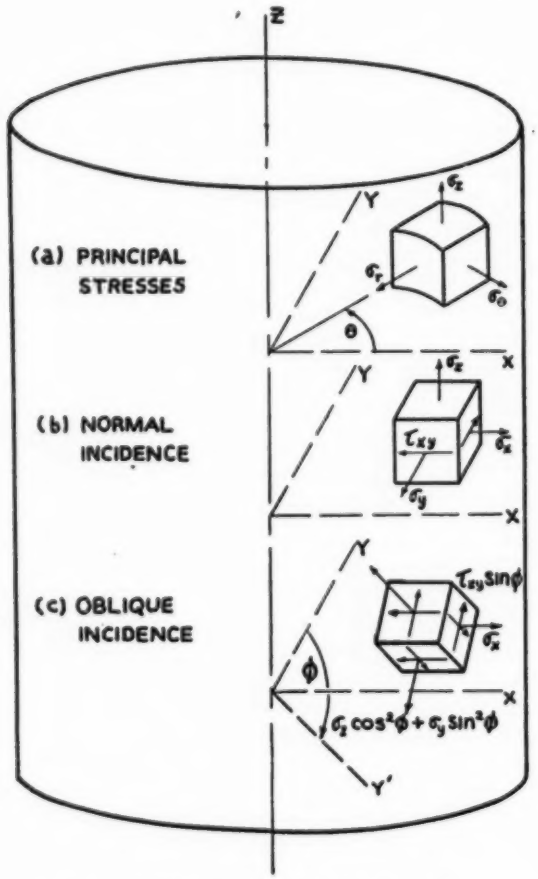


FIG. 1. Axial symmetry. No variation with z.

CIRCULAR CYLINDER

The simplest nontrivial problem to be investigated is the circular cylinder in which the stresses are axi-symmetric and independent of z. Principal stresses are σ_z, σ_r, and σ_θ as shown in Fig. 1. The retardation pattern for parallel light in the y direction is determined by the difference between the secondary principal stresses σ_z and σ_x, Fig. 1(b), where

$$\sigma_x=\sigma_r\cos^2\theta+\sigma_\theta\sin^2\theta=\sigma_\theta+(\sigma_r-\sigma_\theta)\cos^2\theta$$
$$\sigma_z-\sigma_x=(\sigma_z-\sigma_\theta)-(\sigma_r-\sigma_\theta)\cos^2\theta. \quad (1)$$

* The authors are indebted to the National Science Foundation for a grant to study several fundamental phases of three-dimensional photoelasticity.
† Chairman, Division of Engineering.
‡ Research Assistant, Graduate Division of Applied Mathematics.

[1] R. C. O'Rourke and A. W. Saenz, Quart. Appl. Math. 8, 303–311 (1950).
[2] K. A. Parsons, J. Appl. Phys. 24, 469–472 (1953).

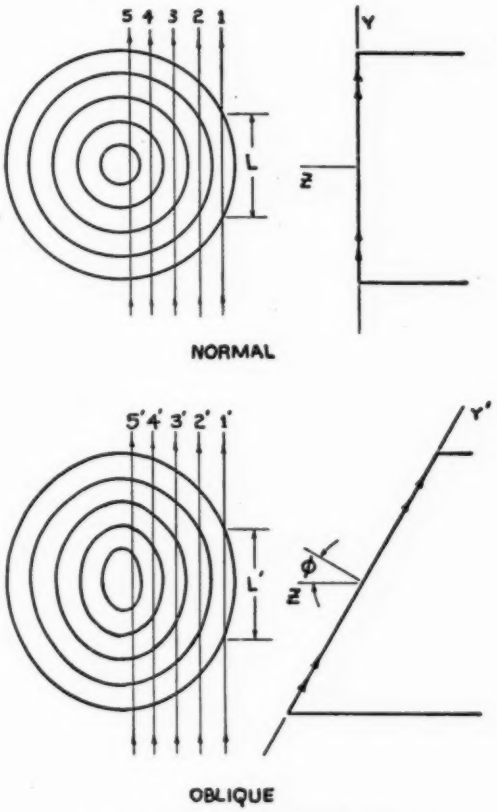


FIG. 2. Normal and oblique incidence.

Designating the retardation in fringe orders by N_y for the path length L

$$N_y=\frac{1}{C}\int_L[(\sigma_z-\sigma_\theta)-(\sigma_r-\sigma_\theta)\cos^2\theta]dy, \qquad (2)$$

where C is the stress-optical coefficient in fringe units for tension or compression.

If light is sent through the model obliquely, at an angle φ to the normal incidence direction, Fig. 1(c), the secondary principal stresses are obtained from the components $\sigma_z\cos^2\varphi+\sigma_y\sin^2\varphi$, $\tau_{zy}\sin\varphi$, and σ_z, where $\tau_{zy}=(\sigma_r-\sigma_\theta)\sin\theta\cos\theta$ and $\sigma_y=\sigma_r\sin^2\theta+\sigma_\theta\cos^2\theta$. The difference between the secondary principal stresses is

$$P-Q=[(\sigma_z\cos^2\varphi+\sigma_y\sin^2\varphi-\sigma_r\cos^2\theta-\sigma_\theta\sin^2\theta)^2 +4(\sigma_r-\sigma_\theta)^2\sin^2\theta\cos^2\theta\sin^2\varphi]^{\frac{1}{2}} \qquad (3)$$

which for $\varphi=45°$ may be written as

$$P-Q=\tfrac{1}{2}\left[(\sigma_z-\sigma_\theta)^2-(\sigma_z-\sigma_\theta)(\sigma_r-\sigma_\theta)(1+3\cos2\theta) +(\sigma_r-\sigma_\theta)^2\left(\frac{3+\cos2\theta}{2}\right)^2\right]^{\frac{1}{2}}. \qquad (4)$$

The retardation in fringes may be expressed as

$$N_0=\frac{1}{C}\int_{L'}(P-Q)dy'. \qquad (5)$$

In principle, the integral expressions for N_y and N_0 are an infinite set of simultaneous equations for $\sigma_z-\sigma_\theta$, and $\sigma_r-\sigma_\theta$ at each of the infinitely many values of the radius r. The principal stress differences, which are the photoelastic unknowns can, therefore, be determined by purely photoelastic means. Strictly speaking, what is found by this technique are the differences between the principal indices of refraction. These in turn are converted to stress differences using a linear relation. Should this relation be invalid at any of the points because of flow, the error will not carry over to neighboring points. Subsequent application of the equation of equilibrium

$$\frac{\partial\sigma_r}{\partial r}+\frac{\sigma_r-\sigma_\theta}{r}=0 \qquad (6)$$

gives σ_r in terms of the value of σ_r at the boundary. The boundary value is zero for most thermal problems. Having σ_r, σ_θ is determined from $\sigma_r-\sigma_\theta$, and then σ_z is found from $\sigma_z-\sigma_\theta$.

In practice, the integral expressions for N_y and N_0 will be replaced by sums over, say, 20 equally spaced intervals of r. As illustrated in Fig. 2, where five rings only are drawn for clarity, a scheme which avoids the difficulty of elaborate simultaneous solution is to take the path 1 nearest the outer radius first. The value of $\cos^2\theta$ is close to unity so that N_y is closely $[(\sigma_z-\sigma_\theta)-(\sigma_r-\sigma_\theta)]L/C$ which determines $\sigma_r-\sigma_\theta$ in terms of $\sigma_z-\sigma_\theta$. Substitution in N_0 for the same location of path, but for oblique incidence, gives $\sigma_z-\sigma_\theta$ and, therefore, $\sigma_r-\sigma_\theta$ for the outer ring. Moving in to the next ring 2 and taking the value of θ at points where path 2 cuts the circular center line of each ring, then to 3, 4, 5, successively, gives all values. If desired, greater accuracy can be achieved by a recalculation using improved estimates for the average values through each ring and taking θ into account more carefully.

Earlier use of equilibrium considerations greatly simplifies the preceding problem but, as stated previously, may introduce error. From the free body sketch, when $\sigma_r=0$ at $r=b$, Fig. 3, $\int_L\sigma_z dy=0$ so that Eq. (2) may be replaced by the much simpler expression

$$N_y=\frac{1}{C}\int_L\sigma_z dy. \qquad (7)$$

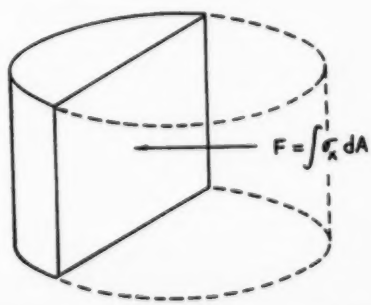


FIG. 3. $\int_L(\sigma_z-\sigma_r)dy=\int_L\sigma_z dy$ when $\sigma_r=0$ on surface of cylinder.

The type of numeric scheme just described now gives σ_z at each of the 20 steps of r from the retardation pattern at normal incidence alone. Oblique incidence gives $\sigma_z-\sigma_\theta$ and, therefore, σ_θ itself at the outside ring, if $\sigma_r=0$ or is known, and enables determination of the interior stresses. An alternative procedure is obtained by following O'Rourke and Saenz and writing the expression for N_y as

$$N_y=\frac{2}{C}\int_{r=z}^{r=b}\frac{\sigma_z r\,dr}{(r^2-x^2)^{\frac{1}{2}}}. \tag{8}$$

This expression may be inverted to give

$$\sigma_z=-\frac{C}{\pi r}\frac{d}{dr}\int_{x=r}^{x=b}\frac{N_y x\,dx}{(x^2-r^2)^{\frac{1}{2}}}=-\frac{C}{\pi}\int_{x=r}^{x=b}\frac{N_y x\,dx}{(x^2-r^2)^{\frac{1}{2}}}$$

$$=-\frac{C}{\pi}\int_{x=r}^{x=b}\frac{\frac{dN_y}{dx}dx}{(x^2-r^2)^{\frac{1}{2}}}, \tag{9}$$

when $\sigma_r=0$ at $r=b$.

Except for notation, this result for σ_z would also have been found by O'Rourke and Saenz if their equation for σ_z had been written explicitly.

Two points might well be emphasized. The first, and most important, is that the inversions of the integrals which correspond to the solution of infinitely many equations with an infinite number of unknowns, have not been found for the more complicated forms. The second point is that the actual evaluation of σ_z at each value of r is not accomplished more easily with the neat integral formulas than with the elementary numeric scheme repeated once to obtain high accuracy.

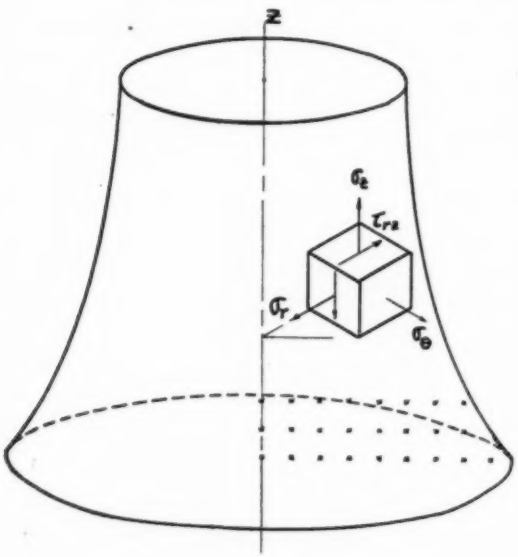


FIG. 4. The axi-symmetric problem. Doubly infinite set of 3 photoelastic unknowns.

AXIAL SYMMETRY AND THE GENERAL CASE

Returning now to the general problem, it is of interest to ignore the computational difficulties and to investigate the possibilities of handling other problems. Summarizing the cylindrical case, there are two photoelastic unknowns along a single infinity of points so that normal and oblique incidence retardation measurements along a corresponding single infinity of points are sufficient to determine them.

In the general axi-symmetric problem there are three photoelastic unknowns at each point of a two dimensional (r,z) infinity of points $\sigma_z-\sigma_\theta$, $\sigma_r-\sigma_\theta$, τ_{rz}, Fig. 4. Each retardation pattern likewise contains a two-dimensional infinity of points so that one normal and two oblique incidence patterns are theoretically sufficient. A possible procedure corresponding to the numeric scheme for the cylindrical case would be to consider equally spaced intervals along both the r and z directions, in effect dividing the model into sectors of three-dimensional annuli with height equal to the interval spacing along z. Tracing the light paths through the model in the normal and two oblique incidence positions would again replace the integral expressions for the retardations with sums over a finite number of intervals and yield a set of simultaneous equations which could be solved for the photoelastic unknowns. As $P-Q$ is given by the square root of a complicated expression, the numeric difficulties should not be under-rated. However, if the photoelastic unknowns are determined, the equations of equilibrium will give the individual stresses.

The general three-dimensional problem is basically in a different class. There are five photoelastic unknowns which may be thought of as $\sigma_x-\sigma_y$, $\sigma_z-\sigma_y$, τ_{xy}, τ_{yz}, τ_{zx} at a three-dimensional infinity of points. As the retardation patterns considered give but a two-dimensional infinitude of information, one normal and four oblique directions of light are required *in infinitely many planes*. Dividing the model into a finite number b^3 of small blocks would seen to indicate that $5b^3$ independent directions of light would suffice. The distinction between the general and the simpler specific cases is, however, that the information is basically absent in the general case. A finer subdivision of the simpler models merely requires a reexamination of the fringe patterns already on hand. Further subdivision of the general three-dimensional requires an enormous number of additional fringe photographs.

The fundamental difficulty is not overcome by the use of the equations of equilibrium, which, while introducing an additional unknown $(\sigma_x+\sigma_y+\sigma_z)$, may be considered to supply but 3 three-dimensional infinitudes of information whereas 5 are required. The deficit cannot be made up by a finite number of two-dimensional patterns. Known values of surface tractions provide at most a two-dimensional infinitude of information. Some much more unusual scheme than now exists must be devised to solve this problem.

A Point Emitter-Junction Collector Transistor*

ROBERT H. KINGSTON
Lincoln Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts
(Received September 10, 1953)

A transistor structure utilizing a planar *p-n* junction as a collector, and a point contact as an emitter has been analyzed theoretically and experimentally. The theory indicates that for maximum frequency cutoff the plane containing the emitter point should be nearly parallel to the collector junction. Also it is found that the base resistance is a critical function of the spacing between the emitter point and the collector junction. Results for experimental models are discussed.

THE point contact and junction transistor have each been studied quite thoroughly over the past few years to the extent that a fairly sound basis of design theory and experiment has been built up to predict the physical and circuit limitations of these devices.[1] It is the purpose of this paper to present pertinent design features of a point emitter-junction collector[2] structure with some tentative experimental results. Such a device has possible advantages over either of the current types since the fabrication requires only a *p-n* junction to which is added a point emitter, thus simplifying the crystal preparation, yet the collector will have a high impedance as in the *n-p-n* or *p-n-p* structure. A possible embodiment of such a transistor is shown in Fig. 1.

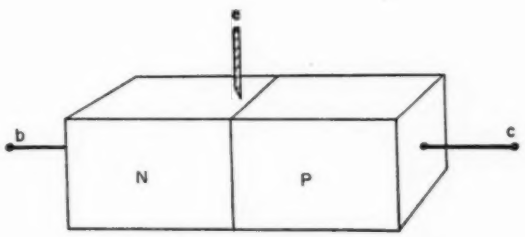


FIG. 1. Prototype point-junction transistor.

DESIGN THEORY

Rather than present a detailed analysis of all the parameters involved in the transistor design the discussion will concentrate on two of the more important device constants since their behavior and physical explanation seem to be unique to this particular structure. These two are the frequency cutoff of α, and the base resistance.

(a) Frequency Cutoff—f_α

Since the collector is a junction, the flow of injected carriers from the emitter is governed completely by diffusion as in the *n-p-n* transistor. A calculation of the frequency cutoff for the diffusion process in the model of Fig. 1 is made by first making the transformations as outlined in Fig. 2. Assuming small signal theory, so that there is no appreciable conductivity-modulation, the flow problem may be represented as in Fig. 2(a), where is shown a point source of minority carriers embedded in a semiconductor at a distance, W, from a plane junction, which is a "sink" for the carriers. This is seen to be the same physical problem if one were to slice through the semiconductor with a plane perpendicular to the junction and containing the emitter point. It does, of course, involve the quite valid assumption that surface recombination may be neglected. The next step is simply to replace the emitter point by an infinite series of points forming an emitter "plane," as in Fig. 2(b). By superposition it is seen, then, that the flow of minority carriers from the plane source will produce the same current at the junction as one point in the plane. The problem is now one-dimensional and may be solved by standard techniques[3] giving a fre-

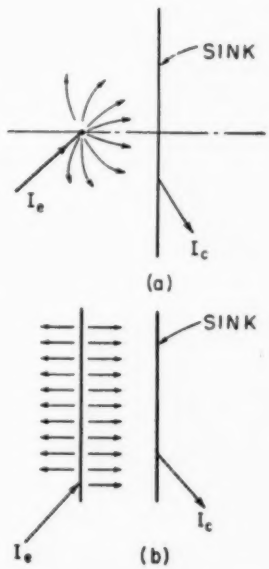


FIG. 2. Calculation of f_α.

* This work was performed by the author while at Bell Telephone Laboratories and portions of it were under Joint Services, Contract No. DA36-039-sc-5589.

[1] See, pp. 1283–1597, e.g., the transistor issue of the Proc. Inst. Radio Engrs. 40, No. 11, 1283–1597 (November, 1952).

[2] W. Shockley, U. S. Patent No. 2,569,347.

[3] W. Shockley, *Electrons and Holes in Semiconductors* (D. van Nostrand Company, Inc., New York, 1950), Chapter 12; also, W. Shockley, Bell System Tech. J. 28, 435 (1949).

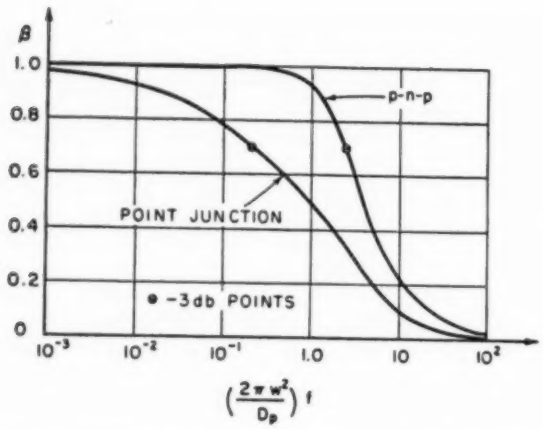


FIG. 3. Calculated variation of $|\beta|$ as a function of frequency.

quency cutoff (defined as the 3 db point or $I_c/I_e=0.707$),

$$f_\alpha=\frac{0.24D_p}{2\pi W^2}, \tag{1}$$

where D_p is the diffusion constant for holes. Now it is interesting to note that this differs from Shockley's result for the *p-n-p* structure,[4]

$$f_\alpha=\frac{D_p}{\pi W^2} \tag{2}$$

by a factor of approximately ten. Examination of Fig. 2(b) shows a marked similarity to the *p-n-p* structure, however, the important difference is the possibility of carriers in the figure escaping to the left of the emitter plane and thus being lost as far as the collector current is concerned. In the junction transistor, however, flow is only allowed to the right of the emitter plane. Neglecting bulk recombination the respective expressions for the β or transmission efficiency from emitter to collector, are

$$\beta=\exp\left[-W\left(i\frac{\omega}{D_p}\right)^{\frac{1}{2}}\right] \tag{3}$$

for the point-junction structure and

$$\beta=\operatorname{sech}W\left(i\frac{\omega}{D_p}\right)^{\frac{1}{2}} \tag{4}$$

for the *p-n-p* structure. These two functions are plotted in Fig. 3, with the respective 3 db points showing the ten-to-one frequency difference. As is seen from the graphs, the beta of the structures is in the ratio of two to one for reasonably high frequencies. This can be explained by realizing that at high frequencies the flow is no longer determined by the proximity of the "sink," and, therefore, for every element of current moving to the right from the emitter plane, the point-junction structure loses one to the left, which is not collected. These considerations suggest a modification of the structure of Fig. 1, to attain higher f_α, which will be discussed below.

(b) Base Resistance

Remembering that the base resistance is defined as the change in open-circuit emitter voltage per unit change in collector current, one is first led to the conclusion that the value of this parameter for the point-junction structure is simply the resistance of the bar from the emitter point to the base contact. This, of course, assumes parallel flow from the collector to the base region, as one might easily conclude. That this point of view is completely erroneous may be demonstrated as a result of the contributions of J. M. Early[5] in connection with the space charge widening at the collector junction. It follows directly from these ideas that, if the collector current increases, this increase is caused, not by a uniform increase in current density across the junction, but by an increase in the effective alpha of the structure, or more specifically an increase of minority carrier flow to the junction in the vicinity of the emitter point. This is shown schematically in Fig. 4(a) where the minority carriers are positive holes. In the diagram, if the collector voltage were increased by a small amount dV, then the collector current would increase, by the Early treatment, as a result of the movement of the space-charge region toward the emitter. Now, when the base resistance is measured, the emitter current is maintained constant, therefore, the increased collector current is supplied from the base contact. It follows that this increase in base current, in reality, is simply a *decrease* in the number of electrons necessary to recombine with holes near the emitter, and thus the collector current

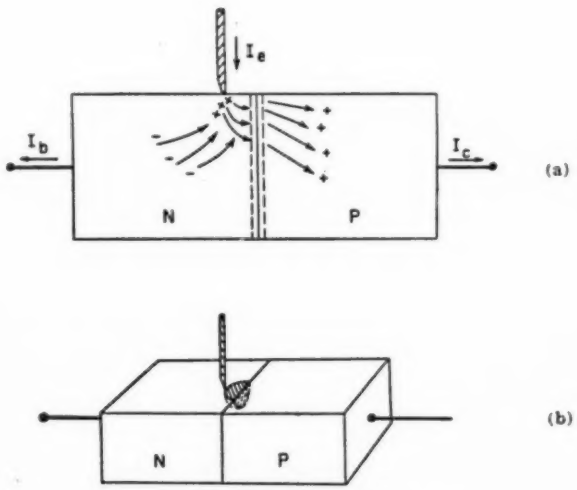


FIG. 4. Base resistance calculation.

[4] W. Shockley, Phys. Rev. 83, 151 (1951).

[5] J. M. Early, Proc. Inst. Radio Engrs. 40, 1401 (1952).

flow, when considering the base resistance, is concentrated in that part of the collector junction near the emitter point. Since the complicated geometry precludes an accurate calculation of this effect, the actual value of the base resistance may be taken to be simply the spreading resistance of the base material from the "active region" near the emitter point. This "active region" will be defined as a quarter of a sphere formed by the collector plane, the top surface of the bar and a radius vector W equal to the distance from the point to the junction. This is the cross-hatched region in Fig. 4(b).

The base resistance is thus seen to be inversely proportional to W, while the frequency cutoff is inversely proportional to W^2. In other words, any increase in frequency cutoff must be paid for by a consequent increase in base resistance.

EXPERIMENTAL RESULTS

As mentioned previously, it becomes advantageous to modify the structure of Fig. 1 to improve the frequency cutoff. This is accomplished as in Fig. 5, where the emitter-collector configuration approaches the form of the *p-n-p* structure, since the injected carriers have much less chance to wander off in a direction away from the junction. Using this design several models were constructed, using a bonded gold point, and base and collector resistivities of the order of 4 ohm-cm. The bars were $\frac{1}{2}$ mm square and 3 mm long. The data of Table I were taken at an emitter current of 0.25 ma and a collector voltage of 1.5 volts. The samples are listed in order of increasing f_α to show the marked concurrent increase in r_b. No attempt was made to optimize alpha which averaged about 0.9 to 0.95. The remaining parameters show a generally good agreement with the complete theory, which follows closely that for the *p-n-p* structure with Early's[5] modifications.

TABLE I. Experimental transistor parameters.

SAMPLE #	①	②	③	④	⑤
$f_{c.o.}$ (Mc)	0.35	0.40	0.48	0.90	1.10
r_b (ohms)	380	340	350	700	1300
W (calculated cm x 10^3)	4.8	4.5	4.1	3.0	2.7
r_e (ohms)	130	250	240	130	160
r_c (megohms)	0.9	2.5	2.5	2.0	2.2
C_c ($\mu\mu$f)	12	13	12	12	9

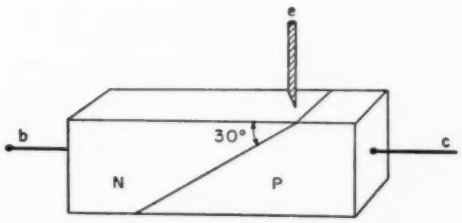


FIG. 5. Modified point-junction structure.

ACKNOWLEDGMENT

The work described is based upon the initial investigations of W. J. Pietenpol of the Bell Telephone Laboratories, who with J. A. Morton directed and encouraged the investigations. The author is also indebted to J. M. Early of the same laboratory for many illuminating discussions of the physical theory.

Concept of Stability for Nuclear Reactors

LAWRENCE BAYLOR ROBINSON*
Atomic Energy Research Department, North American Aviation, Inc., Downey, California
(Received October 5, 1943)

Equations relating neutron density, temperature, and excess multiplication of a nuclear reactor are transformed so as to be suitable for phase plane representation. The concepts of orbital, secular, and asymptotic stability are examined. The significance of each type of stability for reactor operation is discussed. Orbital stability is relevant for reactors whose stable states involve an oscillation of temperature (and neutron density); the other two types involve reactors whose stable states are constant temperatures and neutron density. The latter two types may also be required (for some special considerations) of a reactor possessing orbital stability.

I. INTRODUCTION

ONE of the most important considerations regarding an operating nuclear reactor is the question of its dynamic stability. In this article, the intuitive notion of dynamic reactor stability will be formulated somewhat precisely. General discussions concerning the concept of stability as applied to vacuum tube circuits and mechanical systems have been given by several authors.[1] It is clear that the same stability requirements for a vacuum tube circuit (or a mechanical system) may not be adequate for a reactor.

The characteristics of an operating reactor may be correlated with a system of differential equations; the stability of the reactor is obtained from an examination of the solution of the equations. The mathematical concept of stability is correlated with the physical safety requirements. Several mathematical notions regarding the stability of equations will be examined for relevance in the characterization of reactor operation.

II. DERIVATION OF EQUATIONS

Because of the very strong coupling between neutron density and temperature, it is convenient to consider these as the variables which will yield the required information and to relate the other properties to these two. A simple derivation of the equations to be used follows immediately.[2] Only the essentials will be given so as to display the physical model used. The one group diffusion equation with only prompt neutrons will be the starting point. For a system displaced not too far from equilibrium, this equation is

$$D\nabla^2\phi-\Sigma_a\phi+S=\frac{1}{u}\frac{\partial\phi}{\partial t}. \tag{1}$$

D is the diffusion coefficient (cm), ϕ is the neutron flux (neutrons/cm^2/sec), Σ_a is the macroscopic absorption cross section (cm^{-1}), S is the source term for diffusing (thermal) neutrons, and u is the average speed of the neutrons. A source term of the following type will be used:

$$S=k\Sigma_a\phi\exp(B^2\tau), \tag{2}$$

where k is the multiplication which the given medium would have if it were infinite in extent, B^2 is the critical buckling and τ is the so-called Fermi age of thermal neutrons. When Eq. (2) is substituted into Eq. (1), the following equation is obtained.

$$L^2\nabla^2\phi+[k\exp(-B^2\tau)-1]\phi=\frac{1}{\alpha_0}\frac{\partial\phi}{\partial t}, \tag{3}$$

in which $L^2\equiv D/\Sigma_a$ (the square of the diffusion length) and $\alpha_0\equiv u\Sigma_a$ (the reciprocal of the mean lifetime of thermal neutrons in an infinite medium of the given composition). If one writes ϕ as the product of a space only dependent function and a time dependent part, then

$$L^2\frac{\nabla^2R}{R}+k\exp(-B^2\tau)-1=\frac{1}{\alpha_0 n}\frac{dn}{dt}, \tag{4}$$

where R is the space dependent and n is the time dependent part. Since the system is not far from equilibrium (otherwise it will be inoperative), ∇^2R/R is not very different from $-B^2$ and hence

$$\frac{1}{\alpha_0 n}\frac{1}{1+L^2B^2}\frac{dn}{dt}=\frac{k\exp(-B^2\tau)}{1+L^2B^2}-1, \tag{5}$$

or

$$\frac{1}{\alpha n}\frac{dn}{dt}=k_{eff}-1=k_{ex}. \tag{6}$$

Above $\alpha\equiv\alpha_0(1+L^2B^2)$ is the reciprocal of the mean lifetime of thermal neutrons in the finite medium and $k_{eff}\equiv k\exp(-B^2\tau)(1+L^2B^2)^{-1}$ is the effective multiplication constant of the finite medium. One may partially take into account the effect of the delayed neutrons by using a larger value than α^{-1} for the neutron lifetime, since one of the main consequences of delayed neutrons is an increase in the neutron lifetime

* Now at Naval Research Laboratory (Code 7410), Washington, D. C.

[1] M. J. Kirby, Am. Inst. Elec. Engrs. 69, Part 2, 1662 (1950); F. E. Bothwell, Communication and Electronics, No. 1 (July, 1952); A. A. Andronow and C. E. Charkin, *Theory of Oscillations* (Princeton University Press, Princeton, 1949), pp. 12–15, 22–23, 94–95, 203–208, 319–321.

[2] See S. Glasstone and M. C. Edlund, *The Elements of Nuclear Reactor Theory* (D. van Nostrand Company, Inc., New York, 1952), pp. 290–293. See also J. Chernick, Brookhaven National Laboratory Report-173 (unclassified).

in the medium. The excess of the effective multiplication over unity is given by k_{ex}. It is appropriate to remark at this point that the derivation of Eq. (6) does not depend on the constancy (with respect to time) of k_{eff} (and hence k_{ex}) so long as system remains near enough in the neighborhood of equilibrium so that $\nabla^2 R/R = -B^2$. Thus time variations in k_{ex} are permitted.

The change in multiplication of the medium affects the temperature, which, in turn, affects the multiplication, through the temperature coefficient. The cooling system also tends to remove heat and hence change the temperature. These phenomena are expressed below in quantitative fashion (T will represent temperature).

$$\frac{dn}{dt} = k_{ex}\alpha n, \tag{7}$$

$$k_{ex} = k_{ex}\mathbf{1} - \beta T, \tag{8}$$

$$\frac{dT}{dt} = \gamma n - \lambda T. \tag{9}$$

These are similar to the equations given by J. Chernick.[2] The symbol $\mathbf{1}$ represents Heaviside's unit function and is used to indicate that k_{ex} (which may be time dependent) is suddenly impressed at some t_0. β is the temperature coefficient of multiplication, λ is the relaxation constant for heat removal, and γ is a conversion constant related to the heat released by fission and the specific heat of the medium.

Equations (7), (8), and (9) may be written as the following two equations:

$$\frac{dn}{dt} = k_{ex}\mathbf{1}\alpha n - \beta n T, \tag{10}$$

$$\frac{dT}{dt} = \gamma n - \lambda T. \tag{9}$$

The above is a system of nonlinear first-order differential equations which represents the coupling between neutron density and temperature. It is convenient to eliminate one of the variables and form a single second-order differential equation. The resulting equation is simpler if n, rather than T, is eliminated. This equation is

$$\frac{d^2T}{dt^2} + (\lambda - k_{ex}\mathbf{1}\alpha)\frac{dT}{dt} - k_{ex}\mathbf{1}\alpha\lambda T = -\alpha\beta\left(T\frac{dT}{dt} + \lambda T^2\right). \tag{11}$$

Such an equation is a *nonautonomous* equation because of the possible time dependence of k_{ex}. The solution of such equations is very difficult. This is unfortunate because one would like to set the temperature limits and find out what excursions of k_{ex} can be tolerated. The methods of solution of some nonlinear nonautonomous systems may be found in the work of Stoker, Minorsky, Andronow and Chaikin, and McLachlan.[3]

A transformation, which renders Eq. (11) suitable for so-called phase plane (integral curve) discussions, follows below. The second-order differential equation is converted into two first-order equations. Let

$$T \equiv x_1, \tag{12}$$

$$\frac{dT}{dt} \equiv x_2. \tag{13}$$

It follows that

$$\frac{dx_1}{dt} = x_2 \tag{14}$$

and

$$\frac{dx_2}{dt} = \alpha k_{ex}\mathbf{1}x_1 - (\lambda - \alpha k_{ex}\mathbf{1})x_2 - \alpha\beta(x_1x_2 + \lambda x_1^2). \tag{15}$$

These equations may be written more compactly as

$$\frac{dx_1}{dt} = f_1(x_1, x_2, t), \tag{16}$$

$$\frac{dx_2}{dt} = f_2(x_1, x_2, t), \tag{17}$$

or

$$\frac{d\mathbf{x}}{dt} = \mathbf{f}(\mathbf{x}, t), \tag{18}$$

where

$$\mathbf{x} \equiv (x_1, x_2) \quad \text{and} \quad \mathbf{f} \equiv (f_1, f_2). \tag{19}$$

The vector $\mathbf{x}(t)$ is called a trajectory in phase space. In order that the concept of stability be formulated, it will be assumed that the trajectory $\mathbf{x}(t)$ can be found for any desired conditions. A trajectory, or path, or orbit representing a state of equilibrium will be called the natural or unperturbed motion of the system in question. Any other motions are perturbed motions. P_0 and Q_0 will represent initial points of natural and perturbed motions, respectively; P_t and Q_t are any other points along the respective paths.

One could also work with T and n as x_1 and x_2.

III. TYPES OF STABILITY

A. Orbital Stability

Orbital stability is of importance for periodic motions. Closed trajectories in phase space are indicative of periodic motions. The initial point on such a path will in due time be transformed into itself again. Transformations (or motions) beyond this point will corres-

[3] J. J. Stoker, *Nonlinear Vibrations* (Interscience Publishers, Inc., New York, 1950); N. Minorsky, *Introduction to Nonlinear Mechanics* (Edwards Brothers, Ann Arbor, 1947); A. A. Andronow and C. E. Chaikin, *Theory of Oscillations* (Princeton University Press, Princeton, 1949); N. W. McLachlan, *Ordinary Nonlinear Differential Equations* (Oxford University Press, London, England, 1950).

pond exactly to the previous cycle. A system possessing orbital stability may be said to be stable in the sense of Poincaré, who first gave it precise formulation.[4]

This is a weak form of stability in that additional restraints may be superimposed on it to insure additional forms of stability.

It should be emphasized that this form of stability is applicable only to reactors whose normal operating state involves periodic variations in neutron density and temperature. A transfer to phase space representation will give a closed orbit involving T and dT/dt. A reactor having a given trajectory in phase space such that a small perturbation results in a new trajectory located entirely in the immediate vicinity of the original one, is said to have orbital stability. A stronger requirement could be that a given point of the perturbed system remain arbitrarily close to the point of the natural system from which it arose, rather than that only the trajectories as a whole remain close. For systems having the same period (isochronous), orbital stability implies the stronger type.

B. Secular Stability

Secular stability involves the case in which a point, which is in stable equilibrium, will never wander far away when disturbed. In other words, a small disturbance in a system having secular stability can result only in a small response, which stays small. This type of stability was given precise formulation by Liapounoff.[5] It is seen that the conditions are stronger than those for orbital stability in that the representative points, themselves, of the natural and perturbed motions are required to remain arbitrarily close.

Stability in the sense of Liapounoff may be defined more precisely in the following three equivalent mathematical forms: "A state of equilibrium is stable whenever, given any region ϵ containing it, there is another $\delta(\epsilon)$ in ϵ such that any motion starting in the region δ remains in the region ϵ."[6] Another way of saying this is that if for all disturbed motions satisfying the initial inequality

$$|P_0-Q_0|<\delta(\epsilon), \tag{20}$$

the subsequent motions always satisfy

$$|P_t-Q_t|<\epsilon, \tag{21}$$

the system is stable. Still another statement is that the system is said to possess secular stability if for every small number ϵ, one is able to find a δ (depending on the value of ϵ) such if at time t_0, Q_0 is nearer to P_0 than $\delta(\epsilon)$, then Q_t and P_t are always nearer to each other than ϵ for every $t>t_0$. Above both δ and ϵ are positive.

C. Asymptotic Stability

The requirements for some vacuum tube circuits to possess stability have been given by Nyquist.[7] From one point of view the conditions on this type of stability are stronger than those on secular stability in that the difference between the natural and the perturbed motion is required eventually to vanish. Nyquist states: "The circuit is said to be stable when an impressed small disturbance, which itself dies out, results in a response which dies out. It will be said to be unstable when such a response goes on indefinitely, either staying at a relatively small value or increasing until it is limited by the nonlinearity of the amplifier." The mathematical statement may be given as

$$\lim_{t\to\infty}|P_t-Q_t|=0. \tag{22}$$

It is evident that a nuclear reactor may be stable in one sense and unstable in another. The equations correlating temperature with the other properties of the reactor must be solved in order to learn the type of stability (or instability) the reactor possesses. A change in some physical constants of the reactor as a result of temperature changes may cause a reactor, originally having a form of stability to lose it.

IV. REMARKS

Reactors whose equilibrium states are periodic variations of temperature may have orbital stability as the guiding concept in deciding whether or not they are safe (i.e., suitable for operation in specified location). However, if the difference between the instantaneous point on the natural and one on the perturbed paths is used to actuate a control mechanism, orbital stability may not suffice because this difference can grow to the size of the line connecting the two most widely separated points on the smaller orbit. The actual difference needed to bring the trajectories into coincidence is the distance between a point on the perturbed path and the closest point on the natural path. This latter point will, in general, not be the point which gave rise to the required perturbed point, except for the isochronous case. Hence secular stability (or even asymptotic stability) may be required for some modes of operation.

If the equilibrium point is a constant temperature, then a combination of secular and asymptotic stability more than likely will be required as neither alone may be sufficient. The persistence of a perturbation (as allowed by secular stability) or some wild excursions of a point in phase space (as allowed by asymptotic stability) may cause permanent damage to a reactor. Hence, adequate safety requirements may be had by insisting that the perturbation, which must eventually be out, will never exceed some preassigned value.

[4] H. Poincaré, *New Methods in Celestial Mechanics* (Paris, France, 1892); see also G. G. Birkhoff, *Dynamical Systems* (American Mathematical Society, New York, 1926), pp. 97–122; also Am. J. Math. 49, 1 (1927).

[5] A. M. Liapounoff, "Problème Général de la Stabilité du Mouvement," Annales de la Faculté des Sciences de Toulouse, 1907; also Princeton University Press, 1947, S. Lefshetz, *Lectures on Differential Equations* (Princeton University Press, Princeton, 1948), pp. 72–117.

[6] A. A. Andronow and C. E. Chaikin, reference 1, p. 13.

[7] H. Nyquist, Bell System Tech. J. **11**, 126 (1932).

JOURNAL OF APPLIED PHYSICS VOLUME 25, NUMBER 4 APRIL, 1954

Generalized Off-Axis Distributions from Disk Sources of Radiation*

JACK H. SMITH AND MARTIN L. STORM
Knolls Atomic Power Laboratory,† Schenectady, New York

(Received September 3, 1953)

In the field of shielding it is frequently necessary to calculate biological radiation doses at any point in space due to disk sources of radiation. Such a source is the leakage radiation from an end of a cylindrical reactor. If $G(R)$ is the dose at distance R from a unit point source (e.g., $G\sim e^{-\sigma R}/R^2$) then the dose due to a disk source of radius "a" with source strength $S(\rho)$ is given by

$$D(z,a,\epsilon)=\int_0^a\int_0^{2\pi}G[(z^2+\rho^2+\epsilon^2-2\epsilon\rho\cos\theta)^{\frac{1}{2}}]S(\rho)\rho d\rho d\theta,$$

where (z,ρ,θ) are the usual cylindrical coordinates, $z=0$ is the source plane, and $(z,\epsilon,0)$ locates the dose point.

This integral is quite general, arising in many physical problems. In counter measurements, use of $G=z/(4\pi R^3)$, $S=1$ yields the solid angle subtended by the counter. In neutron physics the slowing down density of neturons from a circular source of finite radius can be calculated in the age approximation with $G\sim\exp(-R^2/\tau)$.

Since for most G's the integral cannot be evaluated in closed form, there are derived three alternative series for $D(z,a,\epsilon)$ each of which has a direct physical interpretation and a different region of convergence. The results are obtained for a class of analytic functions $G(R)$, the possible singularities of which occur only at $R=0$ for finite R; and for $S(\rho)$ of the form $\Sigma_p S_p\rho^{2p}$. The approach in each case is to expand either the double integral or one of the two repeated integrals in a power series in $\cos\theta$, ϵ, or ρ prior to performing the detailed integrations. Numerical results are given for the case of a typical shielding function.

I. INTRODUCTION

IN the field of shielding it is frequently necessary to calculate biological radiation doses (or equivalent detector responses) due to disk sources of radiation. The disks are assumed to be infinitesimally thin so that the sources under consideration are surface distributions. Such a source is the leakage radiation from an end of a cylindrical reactor. The calculation is most easily performed for the case of a small detector located on the axis of the source (Fig. 1). The response $D(z,a)$ is given by

$$D(z,a)=2\pi\int_0^a G[(z^2+\rho^2)^{\frac{1}{2}}]S(\rho)\rho d\rho, \qquad (1)$$

where $G(R)$ is the response of a detector at distance R from a unit source, $z>0$ is the axial distance from disk to detector, "a" is the disk radius, and $S(\rho)$ is the source strength per unit area, assumed axially symmetric. It is to be noted that possible dependence of the source strength per unit area on the angle γ (Fig. 1) is implicitly incorporated in the function $G(R)$ since $\cos\gamma=z/R$ and z is constant in the integration.

When $S(\rho)$ is constant, (1) reduces to the simple form

$$D(z,a)=2\pi S_0\int_z^{R_0}G[\lambda]\lambda d\lambda, \qquad (2)$$

$$R_0=(z^2+a^2)^{\frac{1}{2}}. \qquad (3)$$

Axial symmetry thus facilitates the required summation. However, when it is desired to evaluate the response at a point with any radial coordinate ϵ (Fig. 2), the expression corresponding to (1) is

$$D(z,a,\epsilon)=\int_0^a\int_0^{2\pi}G[(z^2+\rho^2+\epsilon^2-2\epsilon\rho\cos\theta)^{\frac{1}{2}}]\times S(\rho)\rho d\rho d\theta, \qquad (4)$$

where the projection of the line $A'A$ (Fig. 2) on the source plane is taken without any loss in generality as fixing $\theta=0$.

The integral in (4) is quite general, arising in various branches of physics. E.g., if $S(\rho)=1$ and $G(R)=\cos\gamma/(4\pi R^2)$, then $D(z,a,\epsilon)$ is the solid angle subtended by a circular aperture of radius "a" from point A. A survey of formulas for the calculation of this solid angle, often employed in counter measurements, has been given by Jaffey.[1] The result for this special form for G will be examined in Sec. V. In shielding studies G frequently assumes the form

$$G(R)=B(R)(\cos\gamma)^n e^{-\sigma R}/(4\pi R^2), \quad n=0, 1, 2,$$

where $B(R)$ is the "buildup" function and σ is a constant.

In this paper the integration on the right-hand side of (4) is performed for a class of analytic functions $G(R)$, the possible singularities of which occur only at $R=0$ for R finite. The source variations considered are of the form $S(\rho)=\sum_p S_p\rho^{2p}$, $S(\rho)=S_0+S_1\rho^2$ being an adequate description of many sources.

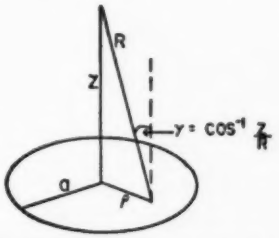


FIG. 1. Geometry for case of axial symmetry.

* This paper is based upon unclassified reports KAPL-918 and KAPL-966 by J. H. Smith and M. L. Storm—Knolls Atomic Power Laboratory.

† Operated by the General Electric Company for the U. S. Atomic Energy Commission.

[1] A. H. Jaffey, Argonne National Laboratory-4875, Sept. 1952. Numerical results are given for a wide range of variables, Rev. Sci. Instr. (to be published).

In Secs. II, III, and IV, three series evaluations of (4) are developed. The leading term of each series is identified as an expression that is used in physical applications as a first approximation to the integral.

The approach in each case is to expand either the double integral in (4) or one of the two single repeated integrals in a power series in $\cos\theta$, ϵ, or ρ prior to performing the detailed integrations. All results can also be obtained formally by expanding $G(R)$ in an appropriate power series, and carrying out termwise integration over both ρ and θ. However, in such a treatment, it is difficult to justify termwise integration by examination of the series for a general G. This difficulty is avoided by the above mentioned approach which, in addition, permits direct determination of the region of convergence of each series.

Equations (14) and (15) in Sec. II are valid for all ϵ, "a," and $z>0$. However, their arithmetical convergence is poor for ϵ/z large. For ϵ/z large but $\epsilon<(z^2+a^2)^{\frac{1}{2}}$ Eqs. (26) and (28) in Sec. III [valid for $\epsilon<(z^2+a^2)^{\frac{1}{2}}$] should be used. Finally, Eq. (40) in Sec. IV is valid for $a<(z^2+\epsilon^2)^{\frac{1}{2}}$ [or $\epsilon>(a^2-z^2)^{\frac{1}{2}}$]. Equations (26), (28), and (40) cover the entire range of parameters. They are simpler in form than (14) and (15) and their use is recommended for numerical applications.

II. ROTATED DISK APPROACH: VALID FOR ALL ϵ, a, AND $z>0$

The first step in this evaluation of the integral in (4) is to note that

$$I=\int_0^a G[(z^2+\rho^2+\epsilon^2-2\epsilon\rho\cos\theta)^{\frac{1}{2}}]S(\rho)\rho d\rho \qquad (5)$$

is an analytic function of $\cos\theta$ for $|\cos\theta|\leqslant 1$. By hypothesis the only singularity of $G(R)$ occurs at $R=0$, but for ρ, ϵ, $z>0$, real and $|\cos\theta|\leqslant 1$ the argument of G is never zero. This guarantees[2] that I is an analytic function of $\cos\theta$, and can be expanded in the following power series:

$$I=\sum_{j=0}^{\infty}\frac{(\cos\theta)^j}{j!}\left\{\frac{\partial^j}{\partial(\cos\theta)^j}\times\int_0^a G[(z^2+\rho^2+\epsilon^2-2\epsilon\rho\cos\theta)^{\frac{1}{2}}]S(\rho)\rho d\rho\right\}_{\cos\theta=0}$$

$$=\sum_{j=0}^{\infty}\frac{(\cos\theta)^j}{j!}\int_0^a(-\epsilon\rho)^j\mathfrak{D}^jG[(z^2+\rho^2+\epsilon^2)^{\frac{1}{2}}]S(\rho)\rho d\rho. \qquad (6)$$

[2] E. T. Whittaker and G. N. Watson, *Modern Analysis*, American Edition, (Macmillan Company, Inc., New York, 1945) p. 92, 5.31.

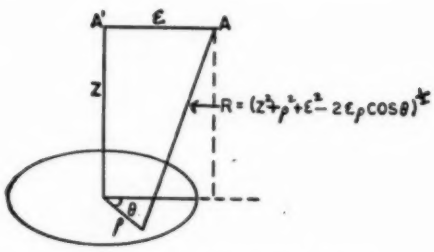


FIG. 2. Geometry for nonaxial case.

In (6) the order of differentiation and integration has been changed and use made of the relation‡

$$\frac{\partial^n}{\partial x^n}G[(\lambda^2+x)^{\frac{1}{2}}]=\frac{1}{2^n}\mathfrak{D}^nG[(\lambda^2+x)^{\frac{1}{2}}], \qquad (7)$$

where the operator $\mathfrak{D}^nG[r]$ is defined by

$$\mathfrak{D}^nG[r]=-\frac{1}{r}\frac{d}{dr}\cdots-\frac{1}{r}\frac{dG}{dr}, \qquad -\frac{1}{r}\frac{d}{dr} \qquad (8)$$

appearing n times.

Substituting (6) into (4) and carrying out the integration over θ termwise (which is possible here) one obtains

$$D(z,a,\epsilon)=2\pi\sum_{j=0}^{\infty}\int_0^a\frac{\epsilon^{2j}\rho^{2j}}{2^{2j}(j!)^2}\times\mathfrak{D}^{2j}G[(z^2+\rho^2+\epsilon^2)^{\frac{1}{2}}]S(\rho)\rho d\rho \qquad (9)$$

terms for odd j vanishing.

Comparison of (1) with the term $j=0$ in (9) shows that the latter represents the response of a detector at A to a virtual source of radius "a." This source is perpendicular to the vector distance R_1 [$R_1=(z^2+\epsilon^2)^{\frac{1}{2}}$] at the intersection of the latter with the actual source (see Fig. 3). The physical interpretation of (9) is that $D(z,a,\epsilon)$ can be calculated by first obtaining the response due to the fictitious disk which lies at the angle $\tan^{-1}\epsilon/z$ with the original, and then modifying the result for the exact effect due to the true source.

It is now convenient to write $S(\rho)$ more explicitly. A physically reasonable form is

$$S(\rho)=\sum_p S_p\rho^{2p}. \qquad (10)$$

Each term from $S(\rho)$ is considered separately and the result of substituting $S_p\rho^{2p}$ into (9) yields D_p:

$$D_p=2\pi S_p\sum_{j=0}^{\infty}\frac{\epsilon^{2j}}{2^{2j}(j!)^2}\times\int_{(z^2+\epsilon^2)^{\frac{1}{2}}}^{(R_0^2+\epsilon^2)^{\frac{1}{2}}}(r^2-z^2-\epsilon^2)^{j+p}\mathfrak{D}^{2j}G[r]rdr, \qquad (11)$$

‡ See Appendix I.

where the change of variable

$$r^2=z^2+\rho^2+\epsilon^2 \tag{12}$$

has been made.

The integration in (11) is done by parts via

$$\int_\alpha^\beta (r^2-d^2)^m \mathfrak{D}^u G[r]rdr$$
$$=\sum_{\sigma=0}^{m}\frac{(-2)^\sigma m!}{(m-\sigma)!}(r^2-d^2)^{m-\sigma}\mathfrak{D}^{u-1-\sigma}G[r]\Big|_\alpha^\beta, \quad m<u$$
$$=(1-\delta_{u0})\sum_{\sigma=0}^{u-1}\frac{(-2)^\sigma m!}{(m-\sigma)!}(r^2-d^2)^{m-\sigma}\mathfrak{D}^{u-1-\sigma}G[r]\Big|_\alpha^\beta \tag{13}$$
$$+\frac{(-2)^u m!}{(m-u)!}\int_\alpha^\beta (r^2-d^2)^{m-u}G[r]rdr, \quad m\geq u.$$

Applying (13), (with $m=j+p$, $u=2j$), to (11) written as

$$\sum_{j=0}^{p}+\sum_{j=p+1}^{\infty}$$

and noting that $j+p\geq 2j$ for $p\geq j$, one obtains for constant source strength $p=0$,

$$D_0=2\pi S_0\int_{(z^2+\epsilon^2)^{\frac{1}{2}}}^{(R_0^2+\epsilon^2)^{\frac{1}{2}}} G[r]rdr+2\pi S_0\sum_{j=1}^{\infty}\frac{\epsilon^{2j}}{2^{2j}j!}$$
$$\times\sum_{\sigma=0}^{j}\frac{(-2)^\sigma}{(j-\sigma)!}(r^2-z^2-\epsilon^2)^{j-\sigma}\mathfrak{D}^{2j-1-\sigma}G[r]\Big|_{(z^2+\epsilon^2)^{\frac{1}{2}}}^{(R_0^2+\epsilon^2)^{\frac{1}{2}}} \tag{14}$$

in which all terms save $\sigma=j$ vanish at the lower limit and $(r^2-z^2-\epsilon^2)=a^2$ at the upper limit.

For variable source strength

$$D_{p\geq 1}=2\pi S_p\sum_{j=0}^{p}\frac{\epsilon^{2j}(j+p)!}{(p-j)!(j!)^2}$$
$$\times\int_{(z^2+\epsilon^2)^{\frac{1}{2}}}^{(R_0^2+\epsilon^2)^{\frac{1}{2}}}(r^2-z^2-\epsilon^2)^{p-j}G[r]rdr$$
$$+2\pi S_p\sum_{j=1}^{\infty}\frac{\epsilon^{2j}}{2^{2j}(j!)^2}\sum_{\sigma=0}^{2j-1}\frac{(-2)^\sigma(j+p)!}{(j+p-\sigma)!}$$
$$\times(r^2-z^2-\epsilon^2)^{j+p-\sigma}\mathfrak{D}^{2j-1-\sigma}G[r]\Big|_{(z^2+\epsilon^2)^{\frac{1}{2}}}^{(R_0^2+\epsilon^2)^{\frac{1}{2}}}. \tag{15}$$

As mentioned earlier in this section, the results obtained are valid for all ϵ, a, $z>0$.

For ϵ/z large the arithmetical convergence of (14) and (15) can be expected to be poor unless $G(R)$ is a slowly varying function of R. This can be seen on physical grounds (see Fig. 3). Important parts of the actual source, as approximated by the virtual disk, are severely rotated out of their natural positions, suggesting that many terms beyond the zero-order approximation are required.

If ϵ/z is large but $\epsilon<(z^2+a^2)^{\frac{1}{2}}$, the results obtained in the next section are more suitable for numerical calculations.

III. DISPLACED DISK APPROACH. VALID FOR $\epsilon<(z^2+a^2)^{\frac{1}{2}}$

In this section

$$\int_0^a\int_0^{2\pi} G[(z^2+\rho^2+\epsilon^2-2\epsilon\rho\cos\theta)^{\frac{1}{2}}]S(\rho)\rho d\rho d\theta$$

will be expanded in a power series in ϵ about $\epsilon=0$. Since the interval of convergence of a power series in a real variable is related directly to the circle of convergence of the same power series with the variable made complex, it will be advantageous to use the analyticity of G with respect to ϵ. By hypothesis $G(R)$ fails to be analytic when and only when $z^2+\rho^2+\epsilon^2-2\epsilon\rho\cos\theta=0$, or

$$\epsilon_1=\rho\cos\theta,$$

and

$$\epsilon_2^2=z^2+\rho^2\sin^2\theta, \tag{16}$$

where $\epsilon=\epsilon_1+i\epsilon_2$.

It is seen in particular that for $|\epsilon|<(z^2+\rho^2)^{\frac{1}{2}}$, $G(R)$ is analytic in ϵ. Since $|\epsilon|$ is constrained to be less than $(z^2+\rho^2)^{\frac{1}{2}}$, and $\rho=0$ appears in the range of integration, it would seem that $|\epsilon|<z$ is required for the double integral $D(z,a,\epsilon)$, to be an analytic function of ϵ. However, the special form of the argument of G is such as to make the result valid for $|\epsilon|<(z^2+a^2)^{\frac{1}{2}}$. This "extension" of the range of validity from $|\epsilon|<z$ to $|\epsilon|<(z^2+a^2)^{\frac{1}{2}}$ is shown in Appendix III.

Thus, for $\epsilon<(z^2+a^2)^{\frac{1}{2}}$ expansion of the double integral into the desired power series yields

$$D(z,a,\epsilon)=\sum_{n=0}^{\infty}\frac{\epsilon^n}{n!}\int_0^a\int_0^{2\pi}\left\{\frac{\partial^n}{\partial\epsilon^n}G[(z^2+\rho^2+\epsilon^2 -2\epsilon\rho\cos\theta)^{\frac{1}{2}}]\right\}_{\epsilon=0}S(\rho)\rho d\rho d\theta. \tag{17}$$

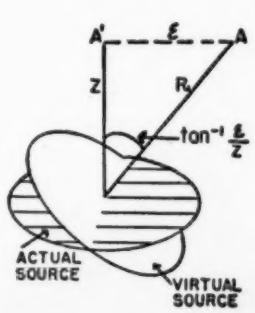


FIG. 3. Actual and virtual sources for rotated disk concept.

Now

$$\frac{1}{n!}\frac{\partial^n}{\partial\epsilon^n}G[(z^2+\rho^2+\epsilon^2-2\epsilon\rho\cos\theta)^{\frac{1}{2}}]$$

$$=\sum_{j=n_0}^{n}\frac{2^{j-n}(\epsilon-\rho\cos\theta)^{2j-n}}{(n-j)!(2j-n)!}$$

$$\times\mathfrak{D}^jG[(z^2+\rho^2+\epsilon^2-2\epsilon\rho\cos\theta)^{\frac{1}{2}}], \quad (18)$$

where

$$n_0=\frac{n+\frac{1}{2}[1-(-1)^n]}{2},$$

as can be verified by induction.

Substitution of (18) into (17) yields

$$D(z,a,\epsilon)=\sum_{n=0}^{\infty}\epsilon^n\int_0^a\int_0^{2\pi}\sum_{j=n_0}^{n}\frac{(-\rho\cos\theta)^{2j-n}2^{j-n}}{(n-j)!(2j-n)!}$$

$$\times\mathfrak{D}^jG[(z^2+\rho^2)^{\frac{1}{2}}]S(\rho)\rho d\rho d\theta. \quad (19)$$

The integration can now be performed termwise over the finite sum. The θ operation leaves only even n's and (19) becomes

$$D(z,a,\epsilon)=\sum_{n=0}^{\infty}2\pi\epsilon^{2n}\sum_{j=n}^{2n}\frac{1}{2^j(2n-j)![(j-n)!]^2}$$

$$\times\int_0^a\mathfrak{D}^jG[(z^2+\rho^2)^{\frac{1}{2}}]\rho^{2(j-n)}S(\rho)\rho d\rho. \quad (20)$$

As in Sec. II consider the term $S_p\rho^{2p}$ from $S(\rho)$. Further, let $k=j-n$ for n fixed and

$$r^2=z^2+\rho^2. \quad (21)$$

Then

$$D_p=2\pi S_p\sum_{n=0}^{\infty}\frac{\epsilon^{2n}}{2^n}\sum_{k=0}^{n}\frac{1}{2^k(n-k)!(k!)^2}$$

$$\times\int_0^a(r^2-z^2)^{p+k}\mathfrak{D}^{k+n}G[r]rdr. \quad (22)$$

Applying (13), (with $m=p+k$, $u=k+n$), to (22) written as

$$\sum_{n=0}^{p}+\sum_{n=p+1}^{\infty}$$

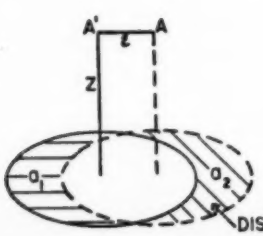


FIG. 4. Actual and displaced sources for displaced disk concept.

and noting that $p+k\geqslant k+n$ for $p\geqslant n$, one obtains

$$D_p=2\pi S_p\sum_{n=p+1}^{\infty}\frac{\epsilon^{2n}}{2^n}\sum_{k=0}^{n}\frac{1}{2^k(n-k)!(k!)^2}$$

$$\times\sum_{\sigma=0}^{k+p}\frac{(-2)^\sigma(k+p)!}{(k+p-\sigma)!}(r^2-z^2)^{p+k-\sigma}\mathfrak{D}^{k+n-1-\sigma}G[r]\Big|_z^{R_0}$$

$$+2\pi S_p\sum_{n=0}^{p}\frac{\epsilon^{2n}}{2^n}\sum_{k=0}^{n}\frac{1}{2^k(n-k)!(k!)^2}$$

$$\times\left\{\frac{(-2)^{k+n}(p+k)!}{(p-n)!}\int_z^{R_0}(r^2-z^2)^{p-n}G[r]rdr\right.$$

$$+(1-\delta_{k+n,0})\sum_{\sigma=0}^{k+n-1}\frac{(-2)^\sigma(p+k)!}{(p+k-\sigma)!}$$

$$\left.\times(r^2-z^2)^{p+k-\sigma}\mathfrak{D}^{k+n-1-\sigma}G[r]\Big|_z^{R_0}\right\}. \quad (23)$$

The finite number of terms in (23) involving integrals are re-expressed as follows:

$$2\pi S_p\sum_{n=0}^{p}\frac{\epsilon^{2n}}{2^n}\sum_{k=0}^{n}\frac{(-2)^{k+n}(p+k)!}{2^k(n-k)!(k!)^2(p-n)!}$$

$$\times\int_z^{R_0}(r^2-z^2)^{p-n}G[r]rdr=2\pi S_p\sum_{n=0}^{p}\frac{\epsilon^{2n}(p!)^2}{[(p-n)!n!]^2}$$

$$\times\int_z^{R_0}(r^2-z^2)^{p-n}G[r]rdr \quad (24)$$

since

$$\sum_{k=0}^{n}\frac{(-1)^k(p+k)!}{(n-k)!(k!)^2}=\frac{(-1)^n(p!)^2}{(p-n)!(n!)^2} \quad (25)$$

as shown in Appendix II. All remaining terms in (23) vanish at the limit $r=z$.

The case $p=0$ (constant source strength) is of special interest and is quite simple:

$$D_0=2\pi S_0\int_z^{R_0}G[r]rdr+2\pi S_0\sum_{n=1}^{\infty}\frac{\epsilon^{2n}}{2^n n}$$

$$\times\sum_{i=1}^{n}\frac{ja^{2j}}{2^j(n-j)!(j!)^2}\mathfrak{D}^{n+j-1}G[R_0]. \quad (26)$$

What was a triple summation in (23) has been reduced to a double summation through manipulations with the finite sums and use of the relation§

$$\sum_{j=v}^{u}\frac{(-1)^j}{(u-j)!j!}=\frac{v}{u}\frac{(-1)^v}{(u-v)!v!}. \quad (27)$$

§ See *Smithsonian Mathematical Formulae and Tables of Elliptic Functions* (1939), p. 19, Sec. 1.51, Eq. (7).

For $p \geq 1$ rearrangement of (23) is straightforward but tedious, yielding

$$D_{p\geq 1}=2\pi S_p \sum_{n=0}^{p} \frac{\epsilon^{2n}(p!)^2}{[(p-n)!n!]^2}\int_a^{R_0}(r^2-z^2)^{p-n}G[r]rdr$$

$$+2\pi S_p(-2)^p \sum_{n=1}^{\infty}\frac{\epsilon^{2n}}{2^n}\sum_{j=p-n+1}^{p+n}\frac{(-1)^j a^{2j}}{2^j j!}$$

$$\times \mathfrak{D}^{n+j-p-1}G[R_0]\sum_{k=j-p}^{n}\frac{(-1)^k(k+p)!}{(n-k)!(k!)^2}. \quad (28)$$

The triple summation in (28) is not so formidable as might appear. First, with the convention that $1/j!=0$ for j a negative integer, many terms in the coefficient of ϵ^{2n} are zero. Secondly, when $(j-p)$ is negative

$$\sum_{k=j-p}^{n}\frac{(-1)^k(k+p)!}{(n-k)!(k!)^2}$$

$$=\sum_{k=0}^{n}\frac{(-1)^k(k+p)!}{(n-k)!(k!)^2}=\frac{(-1)^n(p!)^2}{(p-n)!(n!)^2}, \quad (29)$$

where use has been made of (25). Finally, when $(j-p)$ is positive,

$$\sum_{k=j-p}^{n}\frac{(-1)^k(k+p)!}{(n-k)!(k!)^2}=F(n,p,j) \quad (30)$$

can be evaluated for successive j's via the recursion relation

$$F(n,p,j+1)=F(n,p,j)-\frac{(-1)^{j+p}j!}{(n+p-j)![(j-p)!]^2}, \quad (31)$$

where

$$F(n,p,j=p+1)=\frac{p!}{n!}\left\{\frac{(-1)^n p!}{(p-n)!n!}-1\right\}. \quad (32)$$

The binomial coefficients in (32) are tabulated. Thus the triple sum for D_p is in practice no more than a double sum. Note also that no integral term in (28) is more complicated than the leading one ($n=0$) which is the axial response to a source strength $S_p\rho^{2p}$. A discussion of the physical interpretation of the finite sum of integrals in (28) is given in Appendix IV.

For the case of constant source strength, the particularly simple form of (26) suggests that this result can also be derived by the following physical approach. One considers a fictitious disk (coplanar with the original) the center of which lies a distance ϵ from that of the true source (Fig. 4). By superposition, the response at A due to the original disk is obtained by adding to the effect of the displaced disk that due to area a_1, and subtracting that due to a_2. Thus

$$D_0=2\pi S_0\int_a^{R_0}G(\lambda)\lambda d\lambda$$

$$+S_0\int_{a_1}G(R)d\sigma-S_0\int_{a_2}G(R)d\sigma, \quad (33)$$

where R is the distance from a source point to A and $d\sigma$ is an element of area. The integrations and manipulations hereby implied are lengthy and will not be presented, but the result obtained agrees with (26).

In the next section a third series expansion of (4), convergent for "a"$<(z^2+\epsilon^2)^{\frac{1}{2}}$ [or $\epsilon>(a^2-z^2)^{\frac{1}{2}}$] is derived. Taken together, the formulas of Secs. III and IV provide convergent expansions for the entire range of parameters.

IV. POINT SOURCE APPROACH. VALID FOR $a<(z^2+\epsilon^2)^{\frac{1}{2}}$

In this section the integral in (4) is evaluated by first expanding

$$\int_0^{2\pi}G[(z^2+\rho^2+\epsilon^2-2\epsilon\rho\cos\theta)^{\frac{1}{2}}]d\theta \quad (34)$$

in a power series in ρ. For $\rho=\rho_1+i\rho_2$ the only singularities in G occur for

$$\rho_1=\epsilon\cos\theta, \quad \rho_2^2=z^2+\epsilon^2\sin^2\theta. \quad (35)$$

Thus if $|\rho|<(z^2+\epsilon^2)^{\frac{1}{2}}$, (34) is analytic in ρ. Since ρ takes on the maximum value "a," the required condition for the validity of this development is

$$a<(z^2+\epsilon^2)^{\frac{1}{2}}. \quad (36)$$

Carrying out the desired expansion in (4) one obtains

$$D(z,a,\epsilon)=\int_0^a\sum_{n=0}^{\infty}\frac{\rho^n}{n!}\left\{\frac{\partial^n}{\partial\rho^n}\right.$$

$$\left.\times\int_0^{2\pi}G[(z^2+\rho^2+\epsilon^2-2\epsilon\rho\cos\theta)^{\frac{1}{2}}]d\theta\right\}_{\rho=0}S(\rho)\rho d\rho. \quad (37)$$

Interchanging $\partial^n/\partial\rho^n$ and $\int_0^{2\pi}$

$$D(z,a,\epsilon)=\int_0^a\sum_{n=0}^{\infty}\rho^n\int_0^{2\pi}\sum_{j=n_0}^{n}\frac{(-\epsilon\cos\theta)^{2j-n}2^{j-n}}{(n-j)!(2j-n)!}$$

$$\times\mathfrak{D}^jG[(z^2+\epsilon^2)^{\frac{1}{2}}]d\theta S(\rho)\rho d\rho \quad (38)$$

$$n_0=\frac{n+\frac{1}{2}[1-(-1)^n]}{2},$$

where use has been made of (18) with ρ and ϵ interchanged. Performing the integration in θ over the finite

sum in $\cos\theta$ yields

$$D(z,a,\epsilon)=2\pi\int_0^a\sum_{n=0}^{\infty}\rho^{2n}\times\sum_{j=n}^{2n}\frac{\epsilon^{2(j-n)}\mathfrak{D}^j G[(z^2+\epsilon^2)^{\frac{1}{2}}]}{(2n-j)![(j-n)!]^2 2^j}S(\rho)\rho d\rho. \quad (39)$$

Making the substitution $j-n=k$ for fixed n, again considering $S_p\rho^{2p}$ from $S(\rho)$, and then integrating over ρ one finds

$$D_p=\pi a^2(S_p a^{2p})\sum_{n=0}^{\infty}\frac{a^{2n}}{2^n(n+p+1)}\times\sum_{k=0}^{n}\frac{\epsilon^{2k}}{2^k}\frac{\mathfrak{D}^{k+n}G[(z^2+\epsilon^2)^{\frac{1}{2}}]}{(n-k)!(k!)^2}. \quad (40)$$

The leading term in (40),

$$\frac{\pi a^2(S_p a^{2p})}{p+1}G[(z^2+\epsilon^2)^{\frac{1}{2}}]=\left\{\int_0^{2\pi}\int_0^a S_p\rho^{2p}\rho d\rho d\theta\right\}G[(z^2+\epsilon^2)^{\frac{1}{2}}]$$

is the response of a detector at distance $(z^2+\epsilon^2)^{\frac{1}{2}}$ from a point source—at the center of the disk and of strength equal to the total source strength.

V. APPLICATIONS AND NUMERICAL RESULTS

The results of Secs. III and IV will now be applied to some G's of physical interest. As mentioned earlier $G=z/4\pi R^3$, $S=1$ corresponds to the solid angle

$$\Omega(z,a,\epsilon)=\int_0^a\int_0^{2\pi}\frac{z}{4\pi}\frac{\rho d\rho d\theta}{(z^2+\rho^2+\epsilon^2-2\epsilon\rho\cos\theta)^{\frac{3}{2}}}. \quad (41)$$

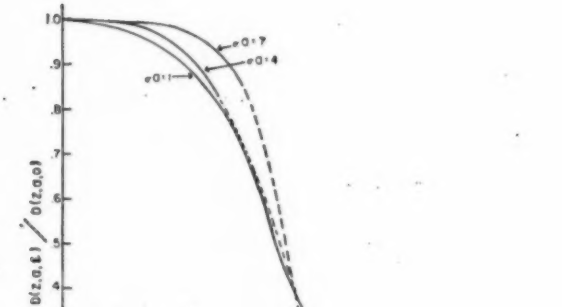


FIG. 5. Variation of $D(z,a,\epsilon)/D(z\,a,0)$ for $z/a=\frac{1}{4}$.

One obtains from (26)

$$\Omega=\frac{1}{2}\left[1-\frac{z}{R_0}\right]-\frac{z}{R_0}\sum_{n=1}^{\infty}\frac{(-1)^n}{n}\left(\frac{\epsilon}{2R_0}\right)^{2n}\times\sum_{m=1}^{n}\frac{(-1)^m m\left(\frac{a}{2R_0}\right)^{2m}[2(n+m)-1]!}{(n-m)!(n+m-1)!(m!)^2} \quad (42)$$

valid for $\epsilon<(z^2+a^2)^{\frac{1}{2}}$.

In the above, use has been made of the relation

$$\mathfrak{D}^s\left(\frac{1}{R^\alpha}\right)=(-1)^s\frac{\alpha(\alpha+2)(\alpha+4)\cdots(\alpha+2s-2)}{R^{2s+\alpha}}. \quad (43)$$

In addition, via (40), the relation valid for $a<(z^2+\epsilon^2)^{\frac{1}{2}}$ is

$$\Omega=\frac{a^2 z}{4(z^2+\epsilon^2)^{\frac{3}{2}}}\sum_{n=0}^{\infty}\left(\frac{a^2}{4[z^2+\epsilon^2]}\right)^n\frac{(-1)^n}{n+1}\times\sum_{k=0}^{n}\left(\frac{\epsilon^2}{4[z^2+\epsilon^2]}\right)^k\frac{(-1)^k(2k+2n+1)!}{(n-k)!(n+k)!(k!)^2}. \quad (44)$$

Equations (42) and (44) permit the calculation of Ω for the entire range of parameters. When written out term by term they are in agreement with Eqs. (10) and (14b) of reference 1.

Numerical studies have been made for the attenuation function

$$G(R)=\frac{e^{-\sigma R}}{4\pi R^2},\quad S(\rho)=S_0. \quad (45)$$

In the foregoing σ is the attenuation coefficient per unit path length.

To proceed, it is necessary first to develop a practical form for the expression

$$\mathfrak{D}^j\left(\frac{e^{-\sigma R}}{R^2}\right).$$

In KAPL 966 there are presented alternative methods for this development. The form of the result which has been employed for numerical work is

$$\mathfrak{D}^j\left(\frac{e^{-\sigma R}}{R^2}\right)=\frac{(-1)^j e^{-\sigma R}}{R^{2j+2}}\sum_{k=0}^{j}(\sigma R)^k\times\sum_{l=0}^{k}\frac{(-1)^l}{(k-l)!l!}(2-l)(4-l)\cdots(2j-l). \quad (46)$$

Although (40) and (26) are now quadruple series, they still are to be preferred over double numerical integrations. This is especially true of (26), for once source dimensions (σa) and distances (R_0) are established, the off-axis variation of D is obtained by calculation of a simple power series in ϵ in which the

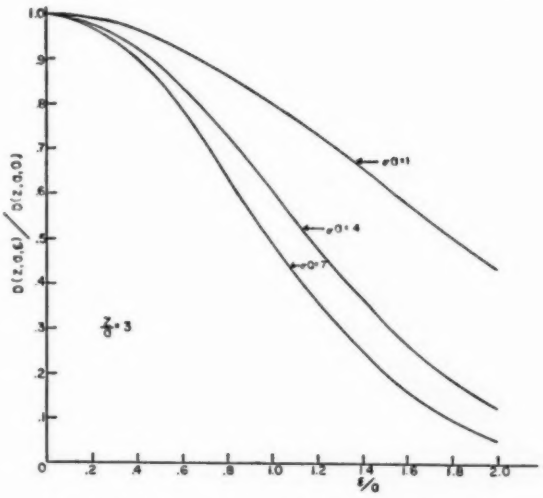


FIG. 6. Variation of $D(z,a,\epsilon)/D(z,a,0)$ for $z/a=3$.

coefficients are numbers for which the determination is straightforward.

Results are presented graphically in Figs. 5–7 for the range $0\leq(\epsilon/a)\leq 2$, $\sigma a=1$, 4, 7, and $z/a=\frac{1}{4}$, 1, 3. All curves are so normalized that $D(z,a,\epsilon=0)=1$, the relative variation in D being of greatest interest here. The shapes of the curves given can be understood quite readily, as follows:

$z/a=\frac{1}{4}$: (see Fig. 5).

For ϵ/a large (≈ 2), the relative dose decreases as σa increases. If a is regarded as a constant, an increasingly good attenuation property of the medium acts to reduce the dose much the same as it would that from a point source at a fixed distance from a receptor. This behavior is typical of the curves for all z/a.

If $\epsilon/a<1$, radiation measured as a function of ϵ/a tends to remain more nearly constant the larger σa. By referring to the discussion of the displaced disk concept in Sec. III, one can see that this is to be expected. The edge corrections which adjust a fictitious dose to an actual dose are rather unimportant, and the effect of an increase in σ is to make them even less important. (The dose point is so close to the source that a "saturation effect" occurs—the major part of the dose originates in a portion of the source which is much smaller than the disk itself).

$z/a=3$: (see Fig. 6).

Near $\epsilon/a=2$, the behavior of D with σa is that already described.

For $\epsilon/a<1$, the trend in D with σa is reversed from that for $z/a=\frac{1}{4}$. At this new increased distance from the source, all parts of it contribute in a more nearly uniform way to the dose. This reversal (D decreasing for ϵ/a fixed and σa increasing) reflects the fact that the two areas involved in the edge correction provide increasingly different doses as the medium attenuates more and more effectively. (The net effect should clearly be a decrease, for it is the radiation from the nearer correction area which is subtracted from the axial dose).

$z/a=1$: (see Fig. 7).

In this intermediate case, the usual effect is observed for large ϵ/a. Near the axis of the system, a good attenuator ($\sigma a=7$) is strong enough to give rise to the "saturation effect," but a moderately good attenuator ($\sigma a=4$) differs from a weak one ($\sigma a=1$) chiefly with regard to the way it transmits radiation from the correction areas a_1 and a_2.

It is relevant now to discuss the accuracy with which relations (26) and (40) can be employed with a reasonable amount of labor. The tables to follow show that six terms in the series are generally sufficient to provide results with an error of about 1 percent (frequently the error is much less). The curves in Figs. 5, 6, and 7 were obtained from tabulations which contained either contributions from enough terms to guarantee about 1 percent accuracy or contributions from six terms. In a few cases, interpolation between accurately known sections of curves was employed. This is illustrated by the dashed portions of the curves for $z/a=\frac{1}{4}$; $\sigma a=4$, 7, and $\epsilon/a\approx 1$.

The following tables give a fairly close indication of the effort required to obtain accuracy of 1 percent or 10 percent. The tabulated numbers represent the numbers of terms required for the accuracy specified. The first term [in $(\epsilon/a)^0$ for (26) and $\epsilon^2(a/\epsilon)^2$ for (40)] is included, and a single term is always characterized by a single index in the leading summation. As is to be expected, numerical convergence is least rapid in regions where the physical approximation implied in the leading term is poorest. For fixed source radius and distance z from the source plane (26) becomes less reliable as ϵ increases. Under the same conditions, (40) becomes less reliable as ϵ decreases toward "a"—a point source approximation being less and less realistic.

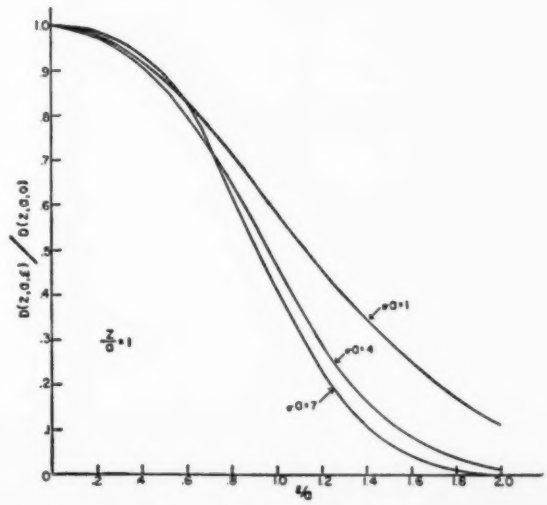


FIG. 7. Variation of $D(z,a,\epsilon)/D(z,a,0)$ for $z/a=1$.

Likewise for ϵ/a and z/a fixed, an increase in σa means that the medium attenuates with increasing effectiveness, and the concomitant variation in the point-source dose function $G(R)$ is difficult to describe with only a few terms of the series.

ACKNOWLEDGMENT

The authors wish to thank Dr. J. Dwork for his continued interest in this problem and the many fruitful discussions held with him. They also wish to acknowledge the help of Dr. G. Roe during the early stages of the work. The calculations were performed by Miss R. Longwell and Miss C. Ranney.

APPENDIX I

If $g(x)=(\lambda^2+x)^{\frac{1}{2}}$, then

$$\frac{\partial^n G[g(x)]}{\partial x^n}=\frac{1}{2^n}\left(\frac{1}{g}\frac{\partial}{\partial g}\right)^n G[g]=\frac{1}{2^n}\mathfrak{D}^n G[g]. \quad \text{(I.1)}$$

This follows by induction. Assuming (I.1) to be true, one finds that

$$\frac{\partial^{n+1}}{\partial x^{n+1}}G[g(x)]=\frac{1}{2^n}\frac{\partial}{\partial x}\left(\frac{1}{g}\frac{\partial}{\partial g}\right)^n G[g]$$

$$=\frac{1}{2^n}\frac{\partial g}{\partial x}\frac{\partial}{\partial g}\left(\frac{1}{g}\frac{\partial}{\partial g}\right)^n G[g]=\frac{1}{2^{n+1}}\left(\frac{1}{g}\frac{\partial}{\partial g}\right)^{n+1} G[g].$$

TABLE I(a). Number of terms required for about 1 percent accuracy: relation No. 26.

σa \ ϵ/a		0.2	0.6	1.0
$z/a=3$	1	1	2	3
	4	2	3	4
	7	2	3	5
$z/a=1$	1	2	2	4
	4	2	3	4
	7	2	3	6
$z/a=\frac{1}{4}$	1	2	3	6
	4	2	5	
	7	3	6	

TABLE I(b). Number of terms required for about 10 percent accuracy: relation No. 26.

σa \ ϵ/a		0.2	0.6	1.0
$z/a=3$	1	1	2	2
	4	1	2	2
	7	1	2	3
$z/a=1$	1	1	2	2
	4	1	3	4
	7	1	3	6
$z/a=\frac{1}{4}$	1	1	2	4
	4	1	3	
	7	1	4	

TABLE II(a). Number of terms required for about 1 percent accuracy: relation No. 40.

σa \ ϵ/a		2.0	1.6	1.2
$z/a=3$	1	3	3	3
	4	3	3	5
	7	4	4	5
$z/a=1$	1	4	4	4
	4	5	5	5
	7	6	6	6
$z/a=\frac{1}{4}$	1	5	5	5
	4	5	5	6

TABLE II(b). Number of terms required for about 10 percent accuracy: relation No. 40.

σa \ ϵ/a		2.0	1.6	1.2
$z/a=3$	1	1	1	2
	4	2	2	2
	7	3	3	3
$z/a=1$	1	3	3	3
	4	3	3	3
	7	4	4	5
$z/a=\frac{1}{4}$	1	3	3	3
	4	4	4	4
	7	6	6	

Since (I.1) is easily seen to be valid for $n=0$ (or 1), it is proved true for any n.

In particular, when $x=0$,

$$\left.\frac{\partial^n}{\partial x^n}G[(\lambda^2+x)^{\frac{1}{2}}]\right|_{x=0}=\frac{1}{2^n}\mathfrak{D}^n G[\lambda]$$

$$=\frac{1}{2^n}\frac{1}{\gamma}\frac{d}{d\gamma}\cdots\frac{1}{\gamma}\frac{dG[\gamma]}{d\gamma}\bigg|_{\gamma=\lambda}.$$

APPENDIX II

$$\sum_{j=0}^{u}\frac{(-1)^j(p+j)!}{(u-j)!(j!)^2}$$

$$=\sum_{j=0}^{u}\frac{(-1)^j}{(u-j)!j!}\frac{d^p}{dt^p}t^{p+j}\bigg|_{t=1}=\frac{d^p}{dt^p}\frac{t^p(1-t)^u}{u!}\bigg|_{t=1}.$$

$$=\sum_{n=0}^{p}\frac{p!}{(p-n)!n!u!}\left[\frac{d^{p-n}}{dt^{p-n}}t^p\right]\left[\frac{d^n}{dt^n}(1-t)^u\right]\bigg|_{t=1}$$

$$=\frac{(-1)^u p!}{(p-u)!u!}\frac{d^{p-u}}{dt^{p-u}}t^p\bigg|_{t=1}=\frac{(-1)^u(p!)^2}{(p-u)!(u!)^2}.$$

APPENDIX III

Consider the integral

$$D(z,a,\epsilon)=\int_0^a\int_0^{2\pi}G[(z^2+\rho^2+\epsilon^2-2\epsilon\rho\cos\theta)^{\frac{1}{2}}]S(\rho)\rho d\rho d\theta. \quad \text{(III.1)}$$

Transforming to rectangular coordinates ($x=\rho\cos\theta$, $y=\rho\sin\theta$) and considering the term $S_p\rho^{2p}$ from $S(\rho)$

$$D_p(z,a,\epsilon)=2S_p\int_{y=0}^{a}\int_{-(a^2-y^2)^{\frac{1}{2}}}^{(a^2-y^2)^{\frac{1}{2}}}G[(z^2+y^2+[x-\epsilon]^2)^{\frac{1}{2}}](x^2+y^2)^p\,dx\,dy \quad \text{(III.2)}$$

or, certainly for $|\epsilon|<z$

$$\frac{\partial D_p}{\partial\epsilon}=2S_p\int_0^a\int_{-(a^2-y^2)^{\frac{1}{2}}}^{(a^2-y^2)^{\frac{1}{2}}}\frac{\partial G[(z^2+y^2+[x-\epsilon]^2)^{\frac{1}{2}}]}{\partial\epsilon}\times(x^2+y^2)^p\,dx\,dy$$
$$=-2S_p\int_0^a\int_{-(a^2-y^2)^{\frac{1}{2}}}^{(a^2-y^2)^{\frac{1}{2}}}\frac{\partial G[(z^2+y^2+[x-\epsilon]^2)^{\frac{1}{2}}]}{\partial x}\times(x^2+y^2)^p\,dx\,dy \quad \text{(III.3)}$$

since $\partial G/\partial\epsilon=-\partial G/\partial x$. Upon integrating by parts there results

$$\frac{\partial D_p}{\partial\epsilon}=-2S_p\int_0^a\{G[(z^2+a^2+\epsilon^2-2\epsilon[a^2-y^2]^{\frac{1}{2}})^{\frac{1}{2}}]-G[(z^2+a^2+\epsilon^2+2\epsilon[a^2-y^2]^{\frac{1}{2}})^{\frac{1}{2}}]\}a^{2p}\,dy$$
$$+2S_p\int_0^a\int_{-(a^2-y^2)^{\frac{1}{2}}}^{(a^2-y^2)^{\frac{1}{2}}}G[(z^2+y^2+[x-\epsilon]^2)^{\frac{1}{2}}]\times\frac{\partial(x^2+y^2)^p}{\partial x}dx\,dy. \quad \text{(III.4)}$$

Consider the arguments of G in the first integral in (III.4), i.e. $z^2+a^2+\epsilon^2\mp2\epsilon[a^2-y^2]^{\frac{1}{2}}$. Letting $\epsilon=\epsilon_1+i\epsilon_2$ it is found that they are equal to zero when and only when

$$\epsilon_1=\pm(a^2-y^2)^{\frac{1}{2}}$$
$$\epsilon_2^2=z^2+y^2 \quad \text{(III.5)}$$

or $|\epsilon|=(z^2+a^2)^{\frac{1}{2}}$. Hence the first integral in (III.4) fails to be an analytic function of ϵ only when $|\epsilon|=(z^2+a^2)^{\frac{1}{2}}$.

This process is performed a total of $2p$ times. Finally, when $\partial^{2p}D_p/\partial\epsilon^{2p}$ is attained, the last term, after integration by parts, is independent of the source factor $(x^2+y^2)^p$ and the result is a finite sum of terms, each analytic in ϵ for $|\epsilon|<(z^2+a^2)^{\frac{1}{2}}$.

Equation (III.2) certainly could have been expanded in a power series ϵ for $|\epsilon|<z$. However, $\partial^{2p}D_p/\partial\epsilon^{2p}$ was shown to have a functional form which is analytic for $|\epsilon|<(z^2+a^2)^{\frac{1}{2}}$, and this derivative could have been obtained from the series for D_p. Since differentiation of a power series does not change its radius of convergence, the series for D_p must also be analytic, not merely for $|\epsilon|<z$, but for $|\epsilon|<(z^2+a^2)^{\frac{1}{2}}$.

APPENDIX IV

Consideration of an infinite plane source provides an understanding of the finite sum of integrals appearing in (32). $D_p(z,a,\epsilon)$ for such a plane with source strength $S_p\rho^{2p}$ can be calculated as follows.

Using a coordinate system centered on the base of the perpendicular from A (Fig. 2) to the plane, obtain D_p' due to a fictitious source of strength $S_p\rho'^{2p}$

$$D_p'=S_p\int_0^\infty\int_0^{2\pi}G[(z^2+\rho'^2)^{\frac{1}{2}}]\rho'^{2p}\rho'\,d\rho'\,d\theta. \quad \text{(IV.1)}$$

Correct this value by the contribution due to the difference between the actual and fictitious sources, both expressed in terms of ρ'. This integral is

$$D_p''=S_p\int_0^\infty\int_0^{2\pi}G[(z^2+\rho'^2)^{\frac{1}{2}}]\times\{(\rho'^2+\epsilon^2+2\epsilon\rho'\cos\theta)^p-\rho'^{2p}\}\rho'\,d\rho'\,d\theta. \quad \text{(IV.2)}$$

Then $D_p=D_p''+D_p'$, or, after carrying out the integrations over θ

$$D_p=2\pi S_p\sum_{n=0}^{p}\frac{\epsilon^{2n}(p!)^2}{[(p-n)!n!]^2}\times\int_0^\infty\rho'^{2(p-n)}G[(z^2+\rho'^2)^{\frac{1}{2}}]\rho'\,d\rho'. \quad \text{(IV.3)}$$

The extension of this concept of superposition to the finite disk would describe the appearance of the finite sum of integrals arising in Sec. III.

Propagation of Longitudinal Deformation Waves in a Prestressed Rod of Material Exhibiting a Strain-Rate Effect*†

ROBERT J. RUBIN
Applied Physics Laboratory, The Johns Hopkins University, Silver Spring, Maryland
(Received September 17, 1953)

We consider the longitudinal propagation of stresses above the yield stress in a material exhibiting a strain-rate effect. The system consists of a semi-infinite rod subjected to end impact. If the rod is prestressed above the yield point and if the impact stress is not too large, the equations of motion are linear. Integral representations of the solutions can be obtained by the method of Laplace transforms. It is shown in general that the strain-rate solutions approach asymptotically the solutions obtained by von Kármán and Taylor in a treatment which neglects the strain-rate effect.

The propagation of the unloading wave is considered for the case in which the initial wave is a shock wave. The condition which defines the regions in the $x-t$ plane, where the loading and unloading equations apply, is given. It is shown that an unloading shock wave will be absorbed except for the case of a material which does not work-harden. In this case the complete solution of the unloading problem is obtained.

INTRODUCTION

WHEN a metal rod is strained in tension past its yield point, one would expect on physical grounds that the speed of a sound pulse in the rod would still be very close to the speed of sound in the unstrained rod. This fact has been established experimentally by Sternglass and Stuart,[1] Bell,[2] Riparbelli,[3] and Campbell.[4] However, from the point of view of static testing, if the stress amplitude is increased and maintained, the rod will simply be elongated to an extent given by the true stress-strain curve for the metal. Thus on the one hand, the rod behaves elastically and on the other hand, flows plastically. It is therefore to be expected that if one considers the propagation of finite amplitude stress pulses in a material strained above its yield point, the front of the pulse will propagate with the sound velocity while the material in the rod flows plastically so long as the stress is maintained at any given cross section.

The experimental work of Bell[2] and Sternglass and Stuart[1] prompted the present analysis. In their experiments, long copper rods were pulled in tension well into the plastic range. While quasi-static tension was gradually being increased, a tensile impact was imparted to a cross section of the rod. The rod was instrumented with strain gauges placed at various distances from the point of impact so that the strains associated with impact could be measured as they propagated down the rod. The results of the experiments showed (1) that the front of the disturbances traveled with sonic velocity and (2) there was a change in the shape of the pulse as it propagated down the rod (whereas, the same pulse in the unstressed rod propagated without dispersion).

The treatment in this paper is intended to provide a qualitative description of the dispersion phenomenon associated with "plastic wave" propagation in a material exhibiting a strain-rate effect.

EQUATIONS OF MOTION FOR A LOADING WAVE

We are concerned with wave propagation in a prestressed rod of material whose "equation of state" is time dependent. It is assumed that elementary cross sections of the rod can support stresses for short periods of time which are greater than the stress which could be supported statically at the same strain. To conform with the experiments of Bell[2] and Sternglass and Stuart,[1] the initial state of tension of the prestressed rod is represented by the point A in Fig. 1. All stresses, strains and displacements are measured from the values corresponding to A. The equations of motion of a load-

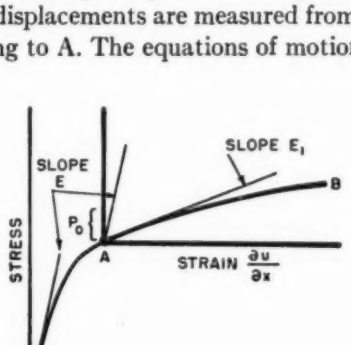


FIG. 1. Typical true stress-strain curve. The point A represents the origin of stress and strain for cross sections of a rod in a prestressed state. The initial reaction of a cross section of the prestressed rod to an increase in stress is elastic (Young's modulus is E).

* This work was supported by the Bureau of Ordnance, Department of the Navy under Contract NOrd 7386.

† Preliminary accounts of this work have appeared: (a) R. J. Rubin, Bull. Am. Phys. Soc. **28**, No. 1, 17 (1953); (b) R. J. Rubin, First Midwestern Conference on Solid Mechanics (1953).

[1] E. J. Sternglass and D. A. Stuart, J. Appl. Mech. **20**, 427 (1953).

[2] J. F. Bell, "Propagation of Plastic Waves in Prestressed Bars," Technical Report No. 5, Navy Contract N6-ONR-243, Task Order VIII, Dept. of Mech. Eng., Johns Hopkins University (1951).

[3] C. Riparbelli, "On the Relation Between Stress, Strain and Velocity in an Element of Bar Submitted to Longitudinal Impact," (1952), Cornell University Aero. Eng. School Report.

[4] W. R. Campbell, J. Wash. Acad. Sci. **43**, 102 (1953); Proc. Soc. Exp. Stress Anal. **10**, (1) 113 (1952).

ing wave in a prestressed rod of material which exhibits a strain-rate effect are[5,6]

$$\rho\frac{\partial^2 u}{\partial t^2}=\frac{\partial P}{\partial x}, \tag{1}$$

$$E\frac{\partial^2 u}{\partial x\partial t}=\frac{\partial P}{\partial t}+\Phi\left[P-F\left(\frac{\partial u}{\partial x}\right)\right], \tag{2}$$

where

$u(x,t)$ = displacement measured from prestressed value.

$P(x,t)$ = stress amplitude measured from the prestressed value.

E = Young's modulus.

ρ = density.

$F(\partial u/\partial x)$ = static stress corresponding to the strain, $\partial u/\partial x$, measured from prestressed values; the initial slope of $F(\partial u/\partial x)$ is the slope of the true stress-strain curve corresponding to the initial prestressed value.

$\Phi[P-F(\partial u/\partial x)]$ = a monotone increasing function which determines the contribution to the rate of strain due to plastic flow where $\Phi[P-F(\partial u/\partial x)]=0$ for $P\leq F(\partial u/\partial x)$.

Equation (1) is merely Newton's law of motion for an elementary cross section and Eq. (2) is the modified "equation of state" for the material. Here the rate of strain, $\partial^2 u/\partial t\partial x$, is proportional to the sum of two terms: the first represents the rate of change of the stress acting at the cross section and the second represents the relaxation process going on in the material, where Φ is an increasing function of the difference $P-F(\partial u/\partial x)$. When the excess stress, $P-F(\partial u/\partial x)$, which acts at a cross section is increased, the rate of the relaxation process will increase. The function $\Phi[P-F(\partial u/\partial x)]=\Phi(z)$ which describes the strain-rate behavior of the material can be expanded in a Maclaurin series about $P-F(\partial u/\partial x)=0$,

$$\Phi\left[P-F\left(\frac{\partial u}{\partial x}\right)\right]=\Phi(0)+\left[P-F\left(\frac{\partial u}{\partial x}\right)\right]\left.\frac{\partial\Phi}{\partial z}\right|_{z=0}+\cdots. \tag{3}$$

The first term on the right side of Eq. (3), $\Phi(0)$, is equal to zero since there is no strain-rate effect when $P=F(\partial u/\partial x)$. We will assume here that for $P-F(\partial u/\partial x)$ not too large

$$\Phi\left[P-F\left(\frac{\partial u}{\partial x}\right)\right]\cong\left[P-F\left(\frac{\partial u}{\partial x}\right)\right]\left.\frac{\partial\Phi}{\partial z}\right|_{z=0}. \tag{4}$$

Malvern and Sokolovsky[5] assumed a $\Phi[P-F(\partial u/\partial x)]$ of the form $k[P-F(\partial u/\partial x)]$ where k may be identified with $\partial\Phi/\partial z|_{z=0}$. It can be expected that for the large stress range which they consider, Eq. (4) will be a poor approximation. Malvern[6] has estimated that k is of the order of 10^{+6} sec^{-1}. From his numerical calculations, he concluded that a strain-rate term of the form in Eq. (4) does not give the correct asymptotic strain at the end of a rod. He found that the asymptotic strain was greater than that predicted by the von Kármán theory.[7] However, we find for the special form of $F(\partial u/\partial x)$ assumed below [see Eq. (5)], that if a constant force is applied to the end of the rod or if the end of the rod is moved at a constant velocity v_0, the strain asymptotically approaches the von Kármán value. Moreover, a proof is outlined in Appendix I showing that even if Malvern's form of $F(\partial u/\partial x)$ is used, the strain must asymptotically approach the von Kármán value, contrary to Malvern's computations.‡

Since we are developing a qualitative description of the Bell, Sternglass and Stuart experiments where tensile impact stress amplitudes are comparatively small (of the order P_0 in Fig. 1), it can be seen that the portion of the stress-strain curve, AB, in Fig. 1 could be reasonably represented by a straight line of slope E_1. Such an assumption results in a great simplification for the subsequent calculations, for the new equations of motion for a loading pulse are linear. Setting $2k=\partial\Phi/\partial z|_{z=0}$, we have

$$\rho\frac{\partial^2 u}{\partial t^2}=\frac{\partial P}{\partial x}, \tag{5}$$

$$E\frac{\partial^2 u}{\partial x\partial t}=\frac{\partial P}{\partial t}+2k\left(P-E_1\frac{\partial u}{\partial x}\right). \tag{6}$$

A principle feature of Eqs. (5) and (6) is that it is possible to obtain an analytic representation of the solution.

PROPAGATION OF LOADING WAVE IN A SEMI-INFINITE ROD

If one considers a tensile loading stress, $P(0,t)$, applied to the end of a semi-infinite rod, the solution of

[5] These equations were first used by V. V. Sokolovsky, Prikladnaia Matematika i Mekhanika 12, 261 (1948) and L. E. Malvern, J. Appl. Mech. 18, 203 (1951) to treat the slightly different problem of propagation of plastic waves in unstressed rods.

[6] The equations of motion change form if the stress falls at any cross section to such a value that $P(x,t)<F(\partial u/\partial x)$. Thereafter an elastic equation of motion applies for the unloading wave.

[7] Th. von Kármán, "On the Propagation of Plastic Deformation in Solids," NDRC Report No. A-29, OSRD No. 365 (1942).

‡ *Note added in proof.*—Malvern (private communication) believes his discrepancy can be attributed to cumulative computational error.

the equations of motion in terms of Laplace transforms is

$$P(x,t)=\frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} g(\sigma) \times\exp\left[t\left\{\sigma-\omega\sigma\left(\frac{\sigma+2k}{\sigma+2k\gamma}\right)^{\frac{1}{2}}\right\}\right]d\sigma, \quad (7)$$

$$u_x(x,t)=\frac{1}{2\pi iE}\int_{c-i\infty}^{c+i\infty} g(\sigma)\frac{\sigma+2k}{\sigma+2k\gamma} \times\exp\left[t\left\{\sigma-\omega\sigma\left(\frac{\sigma+2k}{\sigma+2k\gamma}\right)^{\frac{1}{2}}\right\}\right]d\sigma, \quad (8)$$

$$u(x,t)=-\frac{1}{a\rho}\frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty}\frac{g(\sigma)}{\sigma}\left(\frac{\sigma+2k}{\sigma+2k\gamma}\right)^{\frac{1}{2}} \times\exp\left[t\left\{\sigma-\omega\sigma\left(\frac{\sigma+2k}{\sigma+2k\gamma}\right)^{\frac{1}{2}}\right\}\right]d\sigma, \quad (9)$$

$$u_t(x,t)=-\frac{1}{a\rho}\frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} g(\sigma)\left(\frac{\sigma+2k}{\sigma+2k\gamma}\right)^{\frac{1}{2}} \times\exp\left[t\left\{\sigma-\omega\sigma\left(\frac{\sigma+2k}{\sigma+2k\gamma}\right)^{\frac{1}{2}}\right\}\right]d\sigma, \quad (10)$$

where

$$\omega=\frac{x}{at},$$

$$a=(E/\rho)^{\frac{1}{2}},$$

$$\gamma=E_1/E,$$

and

$$g(\sigma)=\int_0^\infty e^{-\sigma t}P(0,t)dt.$$

The general inversion of Eqs. (7), (8), (9), and (10) has not been possible. However, we will obtain asymptotic solutions for large t as well as explicit solutions for the strain at the end of the rod as a function of time. It might be mentioned here that the solution of Eqs. (7)–(10) for the case $\gamma=0$, i.e., a perfectly plastic material, shows that this plastic material has radically different properties from the work-hardening material ($\gamma\neq 0$).

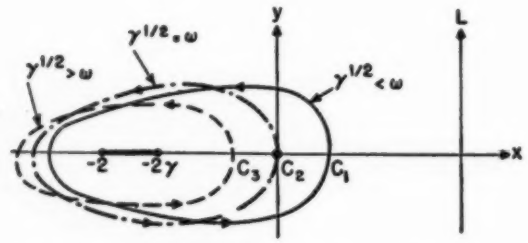


FIG. 2. The three paths of steepest descent for the exponent $\sigma-\omega\sigma(\sigma+2)^{\frac{1}{2}}/(\sigma+2\gamma)^{\frac{1}{2}}$ correspond to the three cases: $\gamma^{\frac{1}{2}}<\omega$, $\gamma^{\frac{1}{2}}=\omega$, and $\gamma^{\frac{1}{2}}>\omega$. The case $\gamma^{\frac{1}{2}}<\omega$ corresponds to that part of the pulse which is near the wave front; $\gamma^{\frac{1}{2}}>\omega$ corresponds to that part of the pulse near the point of impact; and $\gamma^{\frac{1}{2}}=\omega$ corresponds to a critical point in the pulse which separates the regions of small and large plastic strain, and moves at the von Kármán velocity.

ASYMPTOTIC SOLUTIONS

The asymptotic behavior of the solutions of Eqs. (5) and (6) is best illustrated by considering the solutions for the case of a constant stress P_0 applied to the end of the rod at time $t=0$. A typical integral is

$$\frac{E}{P_0}u_x(x,t)=\frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty}\frac{1}{\sigma}\left(\frac{\sigma+2}{\sigma+2\gamma}\right) \times\exp\left[kt\left\{\sigma-\omega\left(\frac{\sigma+2}{\sigma+2\gamma}\right)^{\frac{1}{2}}\right\}\right]d\sigma. \quad (11)$$

If x/a and t are measured in units of k, then the strain is a universal function of the reduced coordinates and γ.

We now proceed to obtain the asymptotic form of Eq. (11) for fixed ω and $kt\to\infty$, using the method of saddle point integration.[8,9] We shall find that the pole at the origin in the integrand of Eq. (11) will, under certain circumstances, coincide with the saddle point. Van der Waerden[9] has developed the method of saddle point integration in a convenient form to handle such coincidences (or near coincidences).

The path of integration L is shown in Fig. 2 in the complex σ plane. There is a cut between $\sigma=-2\gamma$ and $\sigma=-2$. The exponent occurring in the integrand is

$$f(\sigma)=\sigma-\omega\sigma\left(\frac{\sigma+2}{\sigma+2\gamma}\right)^{\frac{1}{2}}. \quad (12)$$

Possible saddle points, σ_0, are defined by the condition that

$$f'(\sigma)=1-\omega\left(\frac{\sigma+2}{\sigma+2\gamma}\right)^{\frac{1}{2}}+\frac{\omega\sigma(1-\gamma)}{(\sigma+2)^{\frac{1}{2}}(\sigma+2\gamma)^{\frac{1}{2}}}=0; \quad (13)$$

and the path of steepest descents passing through the saddle point(s) is given by the condition

$$\mathcal{I}\{f(\sigma)\}=0.$$

It is shown in the Appendix that the path of steepest descents is an oval-shaped curve like C_1, C_2, or C_3 in Fig. 2. There are three different asymptotic forms for Eq. (11) depending on whether the pole at $\sigma=0$ lies inside, on, or outside the closed path of steepest descent. The curves C_1, C_2, and C_3 correspond, respectively, to the following relations between $\gamma^{\frac{1}{2}}$ and ω: $\gamma^{\frac{1}{2}}<\omega$, $\gamma^{\frac{1}{2}}=\omega$, $\gamma^{\frac{1}{2}}>\omega$.

Case I. ($\gamma^{\frac{1}{2}}<\omega$)

The path of integration L can be deformed to the path of steepest descents C_1 and

$$\frac{1}{2\pi i}\int_L(\ )=\frac{1}{2\pi i}\int_{C1}(\ ). \quad (14)$$

[8] P. Debye, Math. Ann. 67, 535 (1909).
[9] B. L. Van der Waerden, Appl. Sci. Research B2, 33 (1950).

Curves of type C_1, where $\sigma=0$ is inside, are characterized by the condition that $\gamma^{\frac{1}{2}}<\omega$. Since $\gamma^{\frac{1}{2}}>1$, the asymptotic formula obtained will apply near the wave front. The saddle point occurs where C_1 crosses the real axis. The asymptotic formula for Eq. (11) for $\gamma^{\frac{1}{2}}<\omega$ and ω constant is

$$u_x(\omega at,t)\sim\frac{P_0}{E}\frac{1}{\sigma_0}\left(\frac{\sigma_0+2}{\sigma_0+2\gamma}\right)^{\frac{1}{2}} \times\frac{\exp\left[kt\left\{\sigma_0-\omega\sigma_0\left(\frac{\sigma_0+2}{\sigma_0+2\gamma}\right)^{\frac{1}{2}}\right\}\right]}{[2\pi ktf''(\sigma)]^{\frac{1}{2}}}. \quad (15)$$

It is shown in the Appendix that

$$\sigma_0-\omega\sigma_0(\sigma_0+2)^{\frac{1}{2}}/(\sigma_0+2\gamma)^{\frac{1}{2}}<0.$$

Thus, $u_x(\omega at,t)$ decreases exponentially to zero for large time near the wave front.

Case II. ($\gamma^{\frac{1}{2}}<\omega$)

When the original path of integration is deformed to the path C_3, the integrals along the two paths are related by

$$\frac{1}{2\pi i}\int_L(\ )=\frac{1}{\gamma}+\frac{1}{2\pi i}\int_{C3}(\ ), \quad (16)$$

where $1/\gamma$ is the residue of the integrand from the pole at $\sigma=0$.

The asymptotic formula for Eq. (11) for $\gamma^{\frac{1}{2}}>\omega$ and ω constant is

$$u_x(\omega at,t)\sim\frac{P_0}{E_1}+\frac{P_0}{E\sigma_0}\left(\frac{\sigma_0+2}{\sigma_0+2\gamma}\right)^{\frac{1}{2}} \times\frac{\exp\left[kt\left\{\sigma_0-\omega\sigma_0\left(\frac{\sigma_0+2}{\sigma_0+2\gamma}\right)^{\frac{1}{2}}\right\}\right]}{[2\pi ktf''(\sigma_0)]^{\frac{1}{2}}}. \quad (17)$$

The quantity $\sigma_0-\omega\sigma_0(\sigma_0+2)^{\frac{1}{2}}/(\sigma_0+2\gamma)^{\frac{1}{2}}$ is again negative. The above asymptotic formula is valid near the impacted end where $u_x(\omega at,t)$ exponentially approaches the von Kármán value, P_0/E_1.

Case III. ($\gamma^{\frac{1}{2}}=\omega$)

For $\gamma^{\frac{1}{2}}=\omega$ the path of steepest descent passes through the pole at $\sigma=0$, the saddle point defined by Eq. (13) being also at $\sigma_0=0$.

Instead of first deforming the path of integration and then making an expansion to obtain the asymptotic formula as previously done, one first introduces a new variable of integration, then deforms the new path of integration, and makes an expansion to obtain the asymptotic formula.

Substituting

$$v=\sigma-\omega\sigma\left(\frac{\sigma+2}{\sigma+2\gamma}\right)^{\frac{1}{2}}=f(\sigma) \quad (18)$$

in Eq. (11), one obtains

$$\frac{E}{P_0}u_x(x,t)=\frac{1}{2\pi i}\int_{L'}\frac{1}{f^{-1}(v)}\left(\frac{f^{-1}(v)+2}{f^{-1}(v)+2\gamma}\right)\frac{\exp(ktv)}{dv/d\sigma}dv, \quad (19)$$

where $\sigma=f^{-1}(v)$. As pointed out by Van der Waerden[9] the saddle point in the σ plane will in general be a branch point in the v plane; the path of steepest descents in the σ plane becomes a series of loops encircling branch points in the v plane such that $|e^{ktv}|$ is as small as possible, i.e., the path of integration is shifted as far to the left as the branch point(s) or pole(s) will permit.

There is a singular point in the integrand of Eq. (19) at $u=0$ due to the presence of the pole at $\sigma=0$ and the saddle point at $\sigma=0$. The order of the singularity can be obtained by expanding $v=f(\sigma)$ in powers of σ and retaining only the lowest order terms. It can easily be shown that for small σ, Eq. (18) becomes

$$v\cong(\tfrac{1}{4}\gamma-\tfrac{1}{4})\sigma^2 \quad (20)$$

and the derivative, $dv/d\sigma$, is

$$\frac{dv}{d\sigma}\cong(\tfrac{1}{2}\gamma-\tfrac{1}{2})\sigma. \quad (21)$$

Substituting Eqs. (20) and (21) in Eq. (19)

$$\frac{E}{P_0}u_x(\gamma^{\frac{1}{2}}at,t)\sim\frac{1}{2\pi i}\frac{1}{2\gamma}\int_C\frac{\exp(ktv)}{v}dv. \quad (22)$$

Thus the two singularities, when combined, yield a simple pole of first order.[10]

The asymptotic behavior of the strain for large time and $\gamma^{\frac{1}{2}}=\omega$, i.e., for a point moving with the velocity $\omega a=(E_1/\rho)^{\frac{1}{2}}$, is

$$u_x(\gamma^{\frac{1}{2}}at,t)\sim\frac{P_0}{2E_1}, \quad (23)$$

a constant which is $\frac{1}{2}$ the asymptotic strain near the impacted end. Recalling the results obtained from the three cases, $\gamma^{\frac{1}{2}}<\omega$, $\gamma^{\frac{1}{2}}>\omega$, and $\gamma^{\frac{1}{2}}=\omega$, it has been found that asymptotically for large t and fixed ω, there is a region of relatively large strain separated from a region of small plastic strain by a transition region which travels with the velocity $(E_1/\rho)^{\frac{1}{2}}$. Thus the front of the region of large strains travels with the velocity assumed by von Kármán[7] and Taylor[11] in theories of

[10] A similar procedure can be used to obtain asymptotic values of $u_x(\gamma^{\frac{1}{2}}at,t)$, $P(\gamma^{\frac{1}{2}}at,t)$, etc.

[11] G. I. Taylor, "The Plastic Wave in a Wire Extended by an Impact Load," British Official Report RC329 (1942).

plastic wave propagation which neglected strain-rate effects.[12]

EFFECT OF A FINITE RISE TIME FOR THE APPLIED STRESS

From a mathematical point of view the presence of a plastic wave front seems to depend on the fact that $g(\sigma)$ in Eq. (8) has a pole of first order at $\sigma=0$. The pole at $\sigma=0$ will be present whenever the stress approaches a constant value after long times. For when the stress is asymptotically $P(0,t)\sim P_0$, then the Laplace transform of $P(0,t)$ is represented asymptotically for small $g(\sigma)\sim P_0/\sigma$.[13] In other words, the behavior of $g(\sigma)$ for small values of σ is determined by the behavior of its inverse transform for large values of the time.

SPECIAL CASE OF SOLUTIONS (7)–(10) FOR CASE OF NO WORK-HARDENING

There is a marked difference in behavior of a work-hardening and a non-work-hardening material as might be expected from the asymptotic formulas obtained in the preceding section. In the work-hardening case, $\gamma\neq 0$, the plastic wave front which develops, travels with tangent-modulus velocity of the plastic part of the stress-strain curve. In the case $\gamma=0$, this tangent-modulus velocity is zero.

When $\gamma=0$, Eqs. (7), (8), and (9) become, for $g(\sigma)=P_0/\sigma$,

$$P(x,t)=\frac{P_0}{2\pi i}\int_{c-i\infty}^{c+i\infty}\exp[t\{\sigma-\omega(\sigma^2+2k\sigma)^{\frac{1}{2}}\}]\frac{d\sigma}{\sigma}, \quad (24)$$

$$u_x(x,t)=\frac{P_0/E}{2\pi i}\int_{c-i\infty}^{c+i\infty}\frac{\sigma+2k}{\sigma^2}\times\exp[t\{\sigma-\omega(\sigma^2+2k\sigma)^{\frac{1}{2}}\}]d\sigma, \quad (25)$$

$$u(x,t)=-\frac{P_0}{(E\rho)^{\frac{1}{2}}}\frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty}\frac{(\sigma+2k)^{\frac{1}{2}}}{\sigma^{b/2}}\exp[t\{\sigma-\omega(\sigma^2+2k\sigma)^{\frac{1}{2}}\}]d\sigma. \quad (26)$$

Using a convolution transformation the expressions for stress, strain, and displacement can be written

$$P(x,t)=P_0\left\{e^{-\xi}+\xi\int_{\xi}^{\tau}e^{-\rho}I_1[(\rho^2-\xi^2)^{\frac{1}{2}}]\times\frac{d\rho}{(\rho^2-\xi^2)^{\frac{1}{2}}}\right\}, \quad \tau>\xi \quad (27)$$

$$u_x(x,t)=\frac{P_0}{E}\left\{(1+2\tau-2\xi)e^{-\xi}+\xi\int_{\xi}^{\tau}(1+2\tau-2\rho)e^{-\rho}\frac{I_1[(\rho^2-\xi^2)^{\frac{1}{2}}]}{(\rho^2-\xi^2)^{\frac{1}{2}}}d\rho\right\}, \quad (28)$$

$$-u(x,t)=\frac{P_0}{k(E\rho)^{\frac{1}{2}}}\int_{\xi}^{\tau}(1+2\tau-2\rho)e^{-\rho}I_0[(\rho^2-\xi^2)^{\frac{1}{2}}]d\rho, \quad (29)$$

where $\xi=kx/a$ and $\tau=kt$. From Eqs. (27) and (28), it can be seen that the stress and strain at the wave front are decreasing exponentially; and asymptotic formulas can be obtained for Eqs. (24) and (25), similar to those obtained in the last section for the region ahead of the plastic wave front.[14] For the present case ($\gamma=0$) there is no analogue of the critical value for the ratio $\gamma^{\frac{1}{2}}/\omega$. In effect, the plastic wave front never moves away from the end of the rod since its velocity should be zero.

The difference in behavior is best illustrated by comparing the displacements and strains of the end of the rod under a suddenly applied constant force, P_0. The asymptotic strain for the cases $\gamma\neq 0$ and $\gamma=0$ are P_0/E_1 and $(P_0/E)2kt$ while the displacements are $-P_0t/(E_1\rho)^{\frac{1}{2}}$ and $-[P_0/(E_1\rho)^{\frac{1}{2}}][(2k)^{\frac{1}{2}}/\Gamma(5/2)]t^{\frac{1}{2}}$, respectively. Thus the asymptotic behavior of the end of the rod in the case of a work-hardening material is identical with that of an elastic material whose Young's modulus is E_1. On the other hand, the asymptotic displacement of the end of the rod for the purely plastic case (no work-hardening) is between the elastic and free-body displacements.

END OF ROD MOVED AT CONSTANT VELOCITY

A simple type of large strain-rate experiment which can be performed is to move the end of the rod at a constant velocity v_0. The solution of Eqs. (5) and (6) for stress and strain with the appropriate initial and boundary conditions are

$$P(x,t)=\frac{Ev_0}{a}\frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty}\frac{1}{\sigma}\left(\frac{\sigma+2\gamma}{\sigma+2}\right)^{\frac{1}{2}}\times\exp\left[kt\left\{\sigma-\omega\sigma\left(\frac{\sigma+2}{\sigma+2\gamma}\right)^{\frac{1}{2}}\right\}\right]d\sigma, \quad (30)$$

$$u_x(x,t)=\frac{v_0}{a}\frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty}\frac{1}{\sigma}\left(\frac{\sigma+2}{\sigma+2\gamma}\right)^{\frac{1}{2}}\times\exp\left[kt\left\{\sigma-\omega\sigma\left(\frac{\sigma+2}{\sigma+2\gamma}\right)^{\frac{1}{2}}\right\}\right]d\sigma. \quad (31)$$

The asymptotic behavior of the integrals in Eqs. (30) and (31) is identical with the behavior of the integrals discussed above in Eqs. (7) and (8). There is a "plastic" wave front which moves with the velocity $(E_1/\rho)^{\frac{1}{2}}$, and

[12] There is one difference, however: E_1 here refers to the slope of the true stress strain curve at the prestressed value of the strain rather than the slope of the engineering stress-strain curve at the prestressed value, as one might naively suppose.

[13] G. Doetsch, *Laplace-Transformation* (Dover Publications, New York, 1943), p. 188.

[14] The path of steepest descents is given in Appendix II.

there is a region of constant strain (behind the wave front) which is equal to the von Kármán value, v_0/a_1. The stress required to move the end of the rod at the velocity v_0 is

$$P(0,t)=\frac{Ev_0}{a}\frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty}\frac{1}{\sigma}\left(\frac{\sigma+2\gamma}{\sigma+2}\right)^{\frac{1}{2}}e^{kt\sigma}d\sigma. \quad (32)$$

Asymptotically $P(0,t)$ is proportional to E_1v_0/a_1. Thus we have an example of a case in which the stress decreases uniformly from its initial value Ev_0/a to E_1v_0/a_1, while the strain increases from its initial value v_0/a to v_0/a_1. It should be noted that the same general behavior is true when $\gamma=0$. When $\gamma=0$, one obtains a result of Malvern,[15] namely,

$$P(0,t)=(Ev_0/a)I_0(kt)\exp(-kt). \quad (33)$$

The asymptotic behavior of Eq. (33) is

$$P(0,t)\sim(Ev_0/a)(2\pi kt)^{-\frac{1}{2}},\quad t\to\infty. \quad (34)$$

The value of the strain at the end is

$$u_x(0,t)=(v_0/a)\exp(-kt)\times[2ktI_1(kt)+(1+2kt)I_0(kt)], \quad (35)$$

and asymptotically the strain at the end is

$$u_x(0,t)\sim(v_0/a)2^{\frac{1}{2}}\pi^{-\frac{1}{2}}(kt)^{\frac{1}{2}}. \quad (36)$$

THE UNLOADING PROBLEM

A treatment of the unloading problem when there is no strain-rate effect has been outlined by E. H. Lee[16] for a nonlinear stress-strain relation. The problem is complicated since there is an irrecoverable strain at each cross section which is "frozen in" so that when the stress level decreases the equation of motion is different. Using the method of characteristics, Lee determined the path of the point in the $x-t$ plane corresponding to the maximum stress level at each cross section for a specimen of finite length. This path divides the $x-t$ plane such that the elastic unloading equation applies on one side and the nonlinear loading equation applies on the other side. In the case of a material exhibiting a strain-rate effect, there is a complication in that the stress, as we have seen, can decrease while the strain will increase at a cross section. In fact, the stress and strain at a cross section can both decrease, but as long as the stress and strain correspond to a point which lies above the static stress strain curve, the rate equation is applicable. Thus one must trace a path, Q, in the $x-t$ plane defined by the condition that $P(x,t)-E_1u_x(x,t)=0$. The curve so obtained corresponds to the curve obtained by Lee. On one side of Q the strain rate equations apply and on the other side of Q the elastic unloading equation applies.

[15] L. E. Malvern, Quart. Appl. Math. **VIII**, 405 (1951).
[16] E. H. Lee, Quart. Appl. Math. **X**, 335 (1953).

If the unloading is assumed to occur abruptly at time, t_0, the curve Q follows the characteristic line $x=a(t-t_0)$ initially and there is a discontinuous change of $P(x,t)$, $u_x(x,t)$, etc., across Q. The change in $P(x,t)$, ΔP, across this unloading shock wave is a measure of its intensity. When the intensity, ΔP, drops to zero, the shock is completely absorbed and the curve Q can then move off the line $x=a(t-t_0)$. The point at which Q can turn off the characteristic can be found by using a result of Malvern[17] that the jump in stress across the unloading shock wave is[18]

$$\Delta P=P_2[a(t-t_0),t]-P_1[a(t-t_0),t]$$
$$=-P_1(0,t_0)+k\int_{t_0}^{t}\{P_1[a(t-t_0),t]-E_1u_{x,1}[a(t-t_0),t]\}dt, \quad (37)$$

where subscript 1 refers to conditions before the shock wave and subscript 2 to conditions after the shock wave has passed.

Inserting Eqs. (7) and (8) in (37) and performing the integration over t, we obtain

$$\Delta P=-\frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty}g(\sigma)e^{t_0\sigma}d\sigma-\frac{1}{4\pi i}\int_{c-i\infty}^{c+i\infty}g(\sigma)\times\left[1+\left(\frac{\sigma+2k}{\sigma+2k\gamma}\right)^{\frac{1}{2}}\right]\left\{\exp\left[t\sigma-(t-t_0)\sigma\times\left(\frac{\sigma+2k}{\sigma+2k\gamma}\right)^{\frac{1}{2}}\right]-\exp(t_0\sigma)\right\}d\sigma. \quad (38)$$

The problem of finding the point at which the curve Q can leave the characteristic $x=a(t-t_0)$ is now reduced to finding the value of t for which the right-hand side of Eq. (38) is zero. Such a calculation will require a numerical solution. However, one may ask the question of whether or not the unloading shock wave will be absorbed. It is a simple matter to show that the shock will always be absorbed.

The quantity $\Delta P(t)$ initially is negative, as may be seen from Eq. (37) where the integral on the right-hand side is zero. If the shock still exists at $t=\infty$, or if it is absorbed there, then $\Delta P(\infty)$ cannot be positive. The expression for $\Delta P(\infty)$ from Eq. (38) is

$$\Delta P(\infty)=\frac{1}{4\pi i}\int_{c-i\infty}^{c+i\infty}g(\sigma)\left[\left(\frac{\sigma+2k}{\sigma+2k\gamma}\right)^{\frac{1}{2}}-1\right]e^{t_0\sigma}d\sigma, \quad (39)$$

where we have made use of the fact that the asymptotic values of the integrals involving t are zero. Equation

[17] L. E. Malvern, "The propagation of longitudinal waves of plastic deformation in a bar of material exhibiting a strain-rate effect," Ph.D. thesis, Brown University (1949).
[18] A derivation of Eq. (37) is given in Appendix III since reference 17 is not readily available.

(38) can be inverted to give

$$\Delta P(\infty)=\frac{1}{2}\int_0^{t_0} P(t_0-\tau)\exp\{-k(1+\gamma)\tau\}\times(I_1[k(1-\gamma)\tau]+I_0[k(1-\gamma)\tau])d\tau. \quad (40)$$

Since we are only dealing with impressed stresses which are positive, we have shown that $\Delta P(\infty)>0$, and hence that the unloading shock was absorbed at some finite time determined from Eq. (38). A numerical calculation is planned to determine whether the unloading shock will undergo internal reflections from the "plastic" wave front in a manner similar to that discussed by Kolsky,[19] for the case in which there is no strain-rate effect.

An implicit assumption was made in Eq. (37) that the condition $P_2(x,t_0)-E_1u_{x,2}(x,t_0)<0$ is satisfied at and near $x=0$, i.e., that the elastic equation of motion governs the motion behind the unloading shock wave near the impacted end of the rod. It is obviously satisfied near $x=0$ since $P_2(0,t_0)=0$ and $u_{x,2}(0,t_0)>0$, for the case $\gamma\neq 0$. However, for the degenerate case $\gamma=0$, the condition $P_2(0,t_0)<0$ is not satisfied, so that the rate equation continues to apply. Consequently, one can obtain the complete solution to the loading-unloading problem for a material which does not work-harden.

For example, if one applies a stress on the end of the rod of the form

$$P(0,t)=\begin{cases} P_0, & 0<t<t_0 \\ 0, & t>t_0 \end{cases} \quad (41a)$$

and

$$g(\sigma)=P_0/\sigma[1-\exp(-\sigma t_0)]. \quad (41b)$$

The solution can be written down immediately. For $0<t<t_0$, Eqs. (27)–(29) apply while for $t>t_0$ behind the unloading wave front, one merely replaces t by $t-t_0$ in the same equations and subtracts [see Eq. 41(b)].

As an example, consider the expression for $P(x,t)$ behind the unloading wave front

$$P(x,t)=P_0\frac{kx}{a}\int_{k(t-t_0)}^{kt} e^{-\rho}I_1\left[\left\{\rho^2-\left(\frac{kx}{a}\right)^2\right\}^{\frac{1}{2}}\right]\times\frac{d\rho}{\left\{\rho^2-\left(\frac{kx}{a}\right)^2\right\}^{\frac{1}{2}}}. \quad (42)$$

Although the complete analytic solution can be obtained quite easily for the case of a material which does not work-harden, the material unfortunately has certain properties which make the application of the solution to the experiments mentioned earlier[1,2] worthless. For example, consider the displacement of the end of the rod as a function of the time for the square pulse of Eq. (41). The displacement is

$$u(0,t)=-P_0(E\rho)^{-\frac{1}{2}}\{t_1F_1(-\tfrac{1}{2};2;-2kt)-(t-t_0)_1F_1(-\tfrac{1}{2};2;-2k(t-t_0))\}; \quad (43)$$

where ${}_1F_1(a;b;x)$ is the confluent hypergeometric function. The asymptotic value of $u(0,t)$ as t goes to infinity is

$$u(0,t)\sim-P_0(E\rho)^{-\frac{1}{2}}\{2^{\frac{1}{2}}(kt)^{\frac{1}{2}}\pi^{-\frac{1}{2}}t_0\}.$$

Thus the displacement of the end of the rod is proportional to the square root of the time. This motion should be compared to the cases of an elastic rod and a free body where the displacements are, respectively, a constant and proportional to the time.

CONCLUSION

It has been shown that for a constant stress suddenly applied to the end of a semi-infinite prestressed rod, the strain is a universal function of a reduced distance and time and a parameter γ. This fact has very interesting experimental consequences. Experiments are now being planned to vary the rate constant k and the slope of the plastic part of the stress-strain curve by using different materials, by using different initial values of the prestress load, and by using different amplitude impact stresses. The aim is to determine whether the present description applies, and if so, to find the range of application. Certainly a strain-rate effect must exist. Whether or not the order of magnitude of the rate constant is 10^{+6} sec^{-1} as estimated by Malvern remains to be seen. *Note added in proof:*—The time constants of the Sternglass and Stuart experiments were such that details of the relaxation process could be lost if the rate constant is of the order 10^{+6} sec.$^{-1}$

ACKNOWLEDGMENTS

The author wishes to thank Dr. R. A. Alpher for helpful discussions on many aspects of this work. He also thanks: Dr. B. A. Fleishman and Dr. R. C. Herman for discussions of the unloading problem; Mrs. Vera Rubin for comments on the manuscript; Mrs. Carol Oberlin for typing the manuscript; and Miss Doris Rubenfeld for the illustrations.

APPENDIX I

Proof that the asymptotic solutions for the strain near the end of a semi-infinite rod obtained from the system of differential equations

$$\rho\frac{\partial^2 u}{\partial t^2}=\frac{\partial P}{\partial x} \quad (A)$$

$$E\frac{\partial^2 u}{\partial x\partial t}=\frac{\partial P}{\partial t}+2k\left[P-F\left(\frac{\partial u}{\partial x}\right)\right] \quad (B)$$

[19] H. Kolsky, *Stress Waves in Solids* (Oxford University Press, London, 1953), p. 173.

for the case of a constant force on the end of the rod or for the case in which the end of the rod is moved at constant velocity, approach monotonically the von Kármán values for the strain in the limit of large time.

Consider the case of a constant stress P_0 applied to the end of the rod. Equation (B), for $x=0$, becomes

$$E\frac{\partial^2 u}{\partial x \partial t}=2k\left[P_0-F\left(\frac{\partial u}{\partial x}\right)\right]. \tag{C}$$

Equation (C) may be integrated to give

$$E\int_0^{\epsilon_t}\frac{d\epsilon}{P_0-F(\epsilon)}=Kt, \tag{D}$$

where $\epsilon=\partial u/\partial x$, ϵ_t is the strain at time t.

If $P_0>F(\epsilon)$ for $\epsilon<\epsilon_0$, where $P_0=F(\epsilon_0)$, then the only way for the integral to increase indefinitely with the time and remain positive is for ϵ_t to approach ϵ_0. Thus the asymptotic strain is that corresponding to the von Kármán value.

Since the asymptotic strain approaches a finite value the rate of strain $\partial^2 u/\partial x \partial t$ asymptotically approaches zero for large time near $x=0$. Inversely, one may conclude from the fact that $\lim_{t\to\infty}\partial^2 u/\partial x\partial t=0$, that the particle velocity at the end of the rod is approaching a constant value, v_0, for large time. Thus if one moves the end of the rod at the constant velocity, v_0, the stress on the end of the rod as a function of time will approach P_0 and the strain will approach ϵ_0 for large time. For it is impossible that two constant asymptotic velocities which are the same, can be produced by different asymptotic stresses. Again these are the von Kármán values.

APPENDIX II. PATH OF STEEPEST DESCENT WHEN $\gamma=0$, $\mathcal{I}\{\sigma-\omega(\sigma^2+2\sigma)^{\frac{1}{2}}\}=0$

When $\gamma=0$ the path of steepest descent defined by the condition, $\mathcal{I}\{f(\sigma)\}=0$, can be obtained quite easily

If $f(\sigma)=\sigma-\omega(\sigma^2+2\sigma)^{\frac{1}{2}}$ and if $\sigma=x+iy$, the condition $\mathcal{I}\{f(\sigma)\}=0$ leads to the following algebraic equation

$$2^{\frac{1}{2}}y-\omega\{[(x^2-y^2+2x)^2+(2y+2xy)^2]^{\frac{1}{2}}-x^2+y^2-2x\}^{\frac{1}{2}}=0. \tag{44}$$

The curve represented by Eq. (44) is an ellipse which intersects the real axis at $\sigma=-1\pm(1-\omega^2)^{-\frac{1}{2}}$.[20] The saddle point(s) will lie on the ellipse at point(s), σ_0 such that

$$f'(\sigma)=1-\omega\frac{\sigma+1}{(\sigma^2+2\sigma)^{\frac{1}{2}}}=0. \tag{45}$$

The roots of Eq. (45) are

$$\sigma_0^{(\pm)}=-1\pm(1-\omega^2)^{-\frac{1}{2}}.$$

[20] The condition, $\mathcal{I}\{f(\sigma)\}=0$, is also satisfied on the real axis for $x<-2$ and for $x>0$.

Thus the saddle points occur where the path of steepest descent crosses the real axis. The value of the function $f(\sigma)$ is greatest at $\sigma_0^{(+)}=-1+(1-\omega^2)^{-\frac{1}{2}}$ and least at $\sigma_0^{(-)}=-1-(1-\omega^2)^{-\frac{1}{2}}$ on the elliptical curve represented by Eq. (44). The value of $f(\sigma)$ at $\sigma_0^{(+)}$ is $f(\sigma_0^{(+)})=-1+(1-\omega^2)^{\frac{1}{2}}$, while $f(\sigma_0^{(-)})=-1-(1+\omega^2)/(1-\omega^2)^{\frac{1}{2}}$. Only $\sigma_0^{(+)}$ is a true saddle point for the integration since $\sigma_0^{(-)}$ does not satisfy the condition that it is a relative maximum on the deformed path of integration. It is quite clear that the original path of the integration can be deformed into the ellipse C which encloses the cut between -2 and 0.

PATH OF STEEPEST DESCENT WHEN $\gamma\neq 0$, $\mathcal{I}\left\{\sigma-\omega\sigma\left(\frac{\sigma+2}{\sigma+2\gamma}\right)^{\frac{1}{2}}\right\}=0$

The form of the path of steepest descent, C, when $\gamma\neq 0$ is best inferred indirectly because the conditions $\mathcal{I}\{f(\sigma)\}=0$ and $f'(\sigma)=0$ lead in general to algebraic equations of the fourth degree, which contain 2 extraneous roots. One can expect that the curve, C, will be similar in form to the ellipse obtained when $\gamma=0$, i.e., an oval-shaped curve symmetric about the real axis which encloses the cut between -2 and -2γ. The function $f(\sigma)=\sigma-\omega\sigma(\sigma+2)/(\sigma+2\gamma)^{\frac{1}{2}}$ is real for real σ except on the cut, and consists of 2 branches separated by the cut. $f(\sigma)$ has the following limiting values: For $\sigma\gg 1$, $f(\sigma)\sim(1-\omega)\sigma\gg 1$; while for $\sigma\ll -1$, $f(\sigma)\sim(1-\omega)\sigma\ll -1$; $\lim_{\sigma\to -2^-}f(\sigma)=-2$, and $\lim_{\sigma\to -2\gamma^+}f(\sigma)=\infty$. Thus, the right branch is infinite at $\sigma=-2\gamma$ and $\sigma=\infty$ and passes through $\sigma=0$. The other point at which $f(\sigma)=0$ can be obtained by setting

$$\sigma-\omega\sigma\left(\frac{\sigma+2}{\sigma+2\gamma}\right)^{\frac{1}{2}}=0. \tag{46}$$

The nonzero root of Eq. (46) is

$$\sigma'=2(\omega^2-\gamma)(1-\omega^2)^{-1}.$$

This root can be positive, negative, or zero, depending on the relative values of ω^2 and γ. Thus for $\omega^2=\gamma$, $\sigma'=0$ and the minimum of the right branch is at the origin. The minimum, $\sigma_{\min}^{(0)}$, in general will lie between $\sigma=0$ and σ' and corresponds to a saddle point. The path of steepest descent will pass through the minimum; and depending on whether the minimum is to the right of the origin, to the left of the origin, or at the origin, the asymptotic formula of the integrals in question will be different. The different asymptotic formulae arise from the fact that the integrands have a pole at the origin which accordingly lies: inside, on, or outside the curve C. The left branch starts at -2, 2 with a negative infinite slope, and therefore passes through a maximum $\sigma_{\max}^{(0)}$ before it goes to minus infinity .It will be shown below that $f(\sigma_{\max}^{(0)})<f(\sigma_{\min}^{(0)})$. We have not as yet excluded the possibility that there are other points in the complex plane which satisfy the conditions $f'(\sigma)=0$

and $\mathcal{I}\{f(\sigma)\}=0$. In order to do this, we proceed as follows. The algebraic equation obtained from $f'(\sigma)=0$ is a quartic containing the parameters ω^2 and γ. In the special case $\gamma=\omega^2=\frac{1}{3}$, the four roots are 0, -2.316, and $-0.841\pm0.244i$. Substitution of the complex roots with $\gamma=\omega^2=\frac{1}{3}$ in

$$f'(\sigma)=1-\omega\left(\frac{\sigma+2}{\sigma+2\gamma}\right)^{\frac{1}{2}}+\frac{\omega(1-\gamma)\sigma}{(\sigma+2\gamma)^2}\left(\frac{\sigma+2\gamma}{\sigma+2}\right)^{\frac{1}{2}} \quad (47)$$

shows that they are extraneous. Since we do not expect any radical changes when $\gamma\neq\omega^2\neq\frac{1}{3}$, we can conclude that there are only two genuine roots of Eq. (47) and that these correspond to $\sigma_{\min}{}^{(0)}$ and $\sigma_{\max}{}^{(0)}$.

One can show that the curve defined by $\mathcal{I}\{f(\sigma)\}=0$ is closed, as follows. For $\sigma=x+yi$, the condition $\mathcal{I}\{f(\sigma)\}$ leads to the equation

$$2^{\frac{1}{2}}y-\omega y\left\{\left[\left(1+\frac{2(1-\gamma)(x+2\gamma)}{(x+2\gamma)^2+y^2}\right)^2+\left(\frac{2(1-\gamma)y}{(x+2\gamma)^2+y^2}\right)^2\right]^{\frac{1}{2}}+1+\frac{2(1-\gamma)(x+2\gamma)}{(x+2\gamma)^2+y^2}\right\}^{\frac{1}{2}}$$

$$+\omega x\left\{\left[\left(1+\frac{2(1-\gamma)(x+2\gamma)}{(x+2\gamma)^2+y^2}\right)^2+\left(\frac{2(1-\gamma)y}{(x+2\gamma)^2+y^2}\right)^2\right]^{\frac{1}{2}}-1-\frac{2(1-\gamma)(x+2\gamma)}{(x+2\gamma)^2+y^2}\right\}^{\frac{1}{2}}=0, \quad (48)$$

where the sign of the last term has the opposite sign of y. Multiplying Eq. (48) by

$$\left\{\left[\left(1+\frac{2(1-\gamma)(x+2\gamma)^2}{(x+2\gamma)^2+y^2}\right)^2+\left(\frac{2(1-\gamma)y}{(x+2\gamma)^2+y^2}\right)^2\right]^{\frac{1}{2}}+1+\frac{2(1-\gamma)(x+2\gamma)}{(x+2\gamma)^2+y^2}\right\}^{\frac{1}{2}}=\Gamma$$

and factoring out y, one obtains

$$2^{\frac{1}{2}}\Gamma-\omega\left[\left(1+\frac{2(1-\gamma)(x+2\gamma)}{(x+2\gamma)^2+y^2}\right)^2+\left(\frac{2(1-\gamma)y}{(x+2\gamma)^2+y^2}\right)^2\right]^{\frac{1}{2}}-\omega-\frac{4\omega(\gamma-\gamma^2)}{(x+2\gamma)^2+y^2}=0. \quad (49)$$

Equation (49) sets bounds on x and y because the terms containing x and y approach zero instead of ω for large x and y. Hence the curve, which is symmetric with respect to the real axis, must be closed. Incidentally, Eq. (49) reduces to Eq. (47) for $y=0$.

We have now reached the point where we know that the path of steepest descent is a closed curve going through $\sigma_{\min}{}^{(0)}$ and $\sigma_{\max}{}^{(0)}$ and that these two points are the only possible saddle points. It merely remains to be shown that $f(\sigma_{\min}{}^{(0)})>f(\sigma_{\max}{}^{(0)})$. Since $f(\sigma_{\min}{}^{(0)})$ is a minimum on the real axis and $f(\sigma)$ is an analytic function at $\sigma_{\min}{}^{(0)}$, $f(\sigma_{\min}{}^{(0)})$ is a maximum on the path of steepest descents which crosses the real axis at right angles and hence $f(\sigma_{\min}{}^{(0)})>f(\sigma_{\max}{}^{(0)})$. Hence there is only one true saddle point for the integration as was shown for the case $\gamma=0$.

APPENDIX III. DERIVATION OF THE RELATION

$$\Delta\mathrm{P}=\mathrm{P}_2[a(t-t_0),t]-\mathrm{P}_1[a(t-t_0),t]=-\mathrm{P}_1(0,t_0)+\mathrm{K}\int_{t_0}^{t}\mathrm{P}_1[a(t-t_0),t]-\mathrm{E}_1\mathrm{u}_{x,1}[a(t-t_0),t]dt$$

Malvern's derivation[17] of the above expression follows naturally from the analysis of the equations of motion (5) and (6) using the method of characteristics. We will here merely assemble the relations necessary for the derivation.

If a stress on the end of the rod suddenly drops to zero at time t_0, then the unloading shock wave moves along the characteristic $x=a(t-t_0)$. A characteristic relation in the unloading region where the elastic wave equation applies is

$$\rho a u_{t,2}-\mathrm{P}_2=\text{constant}. \quad (50)$$

A shock-wave condition which holds across an elastic shock wave is[15]

$$\mathrm{P}_2-\mathrm{P}_1=-\rho a(u_{t,2}-u_{t,1})+\text{constant}. \quad (51)$$

A characteristic relation in the region ahead of the unloading shock is [15]

$$d\mathrm{P}_1-\rho a du_{t,1}=-2k[\mathrm{P}_1-\mathrm{E}_1u_{x,1}]dt. \quad (52)$$

From Eqs. (50) and (51) it follows that

$$\mathrm{P}_2-\mathrm{P}_1=-\tfrac{1}{2}(\mathrm{P}_1-\rho a u_{t,1})+\text{constant}. \quad (53)$$

Taking the differential of (53) along the characteristic $x=a(t-t_0)$ and using (52), one obtains

$$d(\mathrm{P}_2-\mathrm{P}_1)=k[\mathrm{P}_1-\mathrm{E}_1u_{x,1}]dt. \quad (54)$$

Equation (54) may be integrated to yield the desired expression for the jump in stress $\Delta\mathrm{P}$ across the unloading shock wave

$$\Delta\mathrm{P}=-\mathrm{P}_1(0,t_0)+k\int_{t_0}^{t}\mathrm{P}_1[a(t-t_0),t]-\mathrm{E}_1u_{x,1}[a(t-t_0),t]dt. \quad (55)$$

Equation (55) holds for all $t<t^*$ where $\Delta P(t^*)=0$.

Measurements of the Gas Temperature of a Low-Pressure RF Discharge

IRVING REINGOLD AND KENTON GAROFF
Signal Corps Engineering Laboratories, Fort Monmouth, New Jersey
(Received August 3, 1953)

An experimental investigation has been made of the manner in which the gas temperature resulting from an rf sheath discharge is affected by the incident line power, gas fill, and Q of the cavity in which it is confined. Data are presented of the measured gas temperature as a function of line power for cavities of different Q. Qualitative results are submitted of the effectiveness of the gas and of the cavity in dissipating the heat resulting from the power absorbed by the cavity. Argon, hydrogen, nitrogen, and argon and water vapor were the gas fills investigated.

INTRODUCTION

CONSIDERABLE time and effort have been devoted to theoretical and experimental studies of gas discharges. However, very little has been reported in the literature concerning the actual temperature of ionized gases.

This information is of interest and importance in the design and operation of broad-band microwave TR and ATR tubes. These tubes are used in radar systems to connect the antenna and disconnect the receiver from the transmitter during the period of transmission, and to rapidly disconnect the transmitter and connect the receiver to the antenna after transmission. The tube envelope is generally a gas-filled resonant metallic structure with a glass resonant window at the input and output ends in the case of the TR tube, and at the input end in the case of the ATR tube.

An intense rf discharge occurs across the inside face of the input window during the period of transmission. The ability of the glass window to withstand the resultant gas temperature without overheating is often the limiting factor in the system power levels that can be used.

This paper presents the results of an experimental investigation of the temperature of several ionized gases at low pressure confined in resonant ATR cavities of various Q, where the discharge is initiated by a high-power pulse at a frequency of 2820 megacycles. It was the purpose of this investigation to determine by a series of measurements made under controlled laboratory conditions the effect of the characteristics of the gas, Q of the cavity, and the incident rf power on the gas temperature. Spectroscopically pure nitrogen, hydrogen, argon, and argon and water vapor were the gases studied.

EXPERIMENTAL PROCEDURE

A. Conditions Investigated

Each gas studied was confined in an ATR type of cavity fabricated of sheet Kovar, copper brazed to form a quarter-wavelength body of rectangular cross section, and a resonant iris consisting of a rectangular, glass-sealed window. The internal structure of the cavity was nickel-plated. All of the cavities were identical except that windows of different Q were used in their assembly; thus, the difference in the Q of the windows was the only factor that changed the over-all Q of the cavity. Double-iris windows developed at the Signal Corps Engineering Laboratories[1] were incorporated in the quarter-wavelength cavities as well as the more conventional single-iris type of window.

A chromel-alumel thermocouple was sealed into each cavity through the back plate perpendicular to the E field and the glass window, with the thermocouple junction touching the inside surface of the window.

The characteristics of the gas discharge at the inside surface of the input window are similar for both the TR and ATR tubes. However, the fabrication problems, and the incorporation of the thermocouple in the cavity so that the electric field configuration is not disturbed are more straightforward in the ATR than in the TR. Therefore, the ATR type of cavity described was used for this investigation.

Each cavity was baked at a temperature of 400 deg. C for 16 hr while sealed onto a vacuum pumping system which evacuated the cavity to a pressure of 10^{-6} millimeter Hg. The cavity was then filled through liquid air traps with the desired gas to a predetermined pressure. Special techniques were used to insure that the water-vapor filling was free of impurities. Brazing techniques developed at the Signal Corps Engineering Laboratories for the construction of TR and ATR tubes enable the high temperature bakeout process to be used, thus tending to insure that the gas fill will not be contaminated during operation.

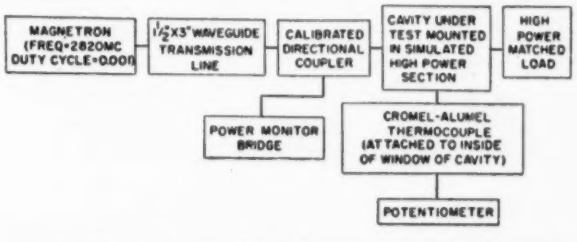

MAGNETRON (FREQ=2820MC DUTY CYCLE=0.001)
½"X3" WAVEGUIDE TRANSMISSION LINE
CALIBRATED DIRECTIONAL COUPLER
CAVITY UNDER TEST MOUNTED IN SIMULATED HIGH POWER SECTION
HIGH POWER MATCHED LOAD
POWER MONITOR BRIDGE
CROMEL-ALUMEL THERMOCOUPLE (ATTACHED TO INSIDE OF WINDOW OF CAVITY)
POTENTIOMETER

FIG. 1. Block diagram showing arrangement of equipment for making gas-temperature measurements.

[1] Reingold, Garoff, and Carter, Proc. Inst. Radio Engrs. 40, 861 (1952).

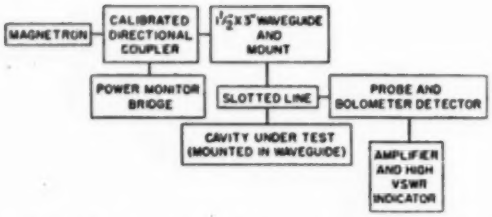


FIG. 2. Block diagram showing arrangement of equipment for making VSWR measurements.

B. Method of Measurement

The equipment used in making the temperature measurements was arranged as shown in Fig. 1.

It was desired to make measurements at peak power levels as high as 2 megawatts. Since the peak power output of the magnetrons available for use was in the order of 0.75 megawatt, a simulated high-power section with an appropriate power stepup ratio was designed. This simulated high-power section is a length of transmission line an integral number of half-wavelengths long in which a very large standing wave ratio is matched to the transmission line by inductive irises, thus forming a high Q resonant cavity.[2] The structure under test is mounted in this transmission line resonant cavity at a high current point which is equivalent to that which would exist at a high line power level.

Measurements of the temperature of the gas discharge at the inside of the window were taken as a function of peak line power up to the equivalent 2-megawatt level. The pulse length was 1 microsecond and the pulse repetition frequency was 1000 pulses per second in each case. Proper precautions were taken to insure that equilibrium conditions were attained at each power level.

A preliminary series of checks were made to insure that the thermocouple did not act as a probe and that the actual gas temperature was not masked by the electron or ion temperatures. The first cavity was built with the thermocouple leads mounted through a flexible diaphragm which enabled the depth of the thermocouple to be varied by over $\frac{1}{16}$ of an in., a distance sufficiently great to remove the tip of the thermocouple from the thin plasma sheath adjacent to and covering the input window. Temperature measurements made while the thermocouple position was varied in this manner showed no evidence that any probing effect was present.

Measurements of VSWR *versus* peak line power were made on each cavity for each gas fill in order to distinguish between the effect of the gas fill and that of the window on the resultant gas temperature. These results were then converted to arc loss as a function of the line power,[3] and the average power absorbed by each cavity for each gas fill was calculated.

The equipment was arranged as shown in Fig. 2.

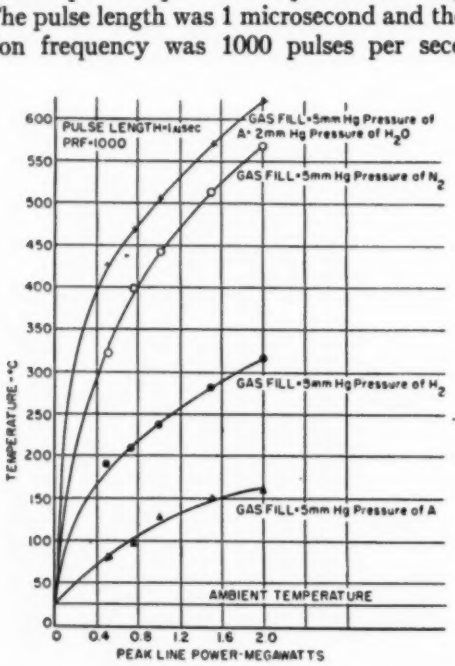

PULSE LENGTH=1 μsec
PRF=1000
GAS FILL=5mm Hg Pressure of A+2mm Hg Pressure of H_2O
GAS FILL=5mm Hg Pressure of N_2
GAS FILL=5mm Hg Pressure of H_2
GAS FILL=5mm Hg Pressure of A
AMBIENT TEMPERATURE
TEMPERATURE - °C
PEAK LINE POWER-MEGAWATTS

FIG. 3. Gas temperature of different gas fills *versus* line peak power for cavity with window $Q=0.6$. Value of window Q is for the doubly-loaded condition.

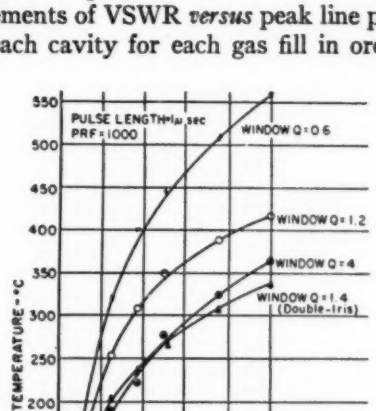


FIG. 4. Gas temperature of nitrogen-gas fill *versus* line peak power for cavities of different window Q. Value of window Q is for the doubly-loaded condition.

EXPERIMENTAL RESULTS

The curves shown in Figs. 3 and 4 are representative of the relationship between the peak line power and the gas temperature for the cavities and the gas fills investigated. As expected, the gas temperature for the same line power varies inversely as the Q of the resonant window in the case of the cavities with the single-iris

[2] Montgomery, Dicke, and Purcell, *Principles of Microwave Circuits* (McGraw-Hill Book Company, Inc., New York, 1948), Radiation Laboratory Series, Vol. 8, pp. 182–186.

[3] Material Laboratory, New York Naval Shipyard, Tech. Report NE091105, Project 5032-K-20.13 (May 28, 1952), pp. 44–46, Enclosure 12.

windows. However, the gas temperatures of the cavity with the double-iris resonant window is considerably lower than that of the cavity with a single-iris window of comparable Q. Note that the use of the argon and water vapor gas fill results in the highest gas temperature, with nitrogen, hydrogen, and argon following in that order. It should also be noted that there is a tendency for the gas temperature to level off with increasing line power.

The temperature at the exposed portion of the tube body at the 2-megawatt peak line power level ranged from 40 degrees C for an argon gas fill in the cavity with the lowest Q window, to 100 degrees C for an argon and water vapor gas fill in the cavity with the highest Q window.

Table I shows the relationship between the peak line power and the average power absorbed by the cavity as a function of the gas fill. The power absorbed varies inversely as the Q of the resonant window of the cavity for each gas fill, but the power absorbed by the cavity

TABLE I. Average power absorbed by a cavity with a doubly-loaded window Q of 0.6 for the gas fills investigated. The data marked * is extrapolated.

	Average power absorbed by cavity			
Peak line power (megawatts)	Gas fill 5 mm Hg pressure of A+2 mm Hg pressure of H_2O (watts)	Gas fill 5 mm Hg pressure of N_2 (watts)	Gas fill 5 mm Hg pressure of H_2 (watts)	Gas fill 5 mm Hg pressure of A (watts)
0.125 megawatt	13 w	11 w	10 w	4 w
0.250 megawatt	20	18	19	7
0.350 megawatt	22	20	22	9
0.500 megawatt	25	25	28	11
0.750 megawatt	32	32	39	13
1.0 megawatt	40*	45*	51*	16*
1.5 megawatts	54*	66*	75*	22*
2.0 megawatts	66*	86*	98*	30*

with the double-iris resonant windows, although reasonably close to that of the cavity with a single-iris window of comparable Q, is somewhat higher in each case.

The curves presented in Fig. 5 show the relationship between the measured gas temperature and the power absorbed by the cavity as a function of the gas fill. These curves are actually a measure of the "effective heat conductivity" of the gas. Note that argon is the coolest gas fill for a given average power absorbed by the cavity, with hydrogen, nitrogen, and argon and water vapor following in that order.

The curves presented in Fig. 6 show the relationship between the measured gas temperature and the power absorbed by each cavity with window Q as a parameter. This set of curves may be considered a measure of the "effective heat conductivity" of the resonant window.

DISCUSSION OF RESULTS

The properties of the glass used for the resonant windows are naturally of importance in determining

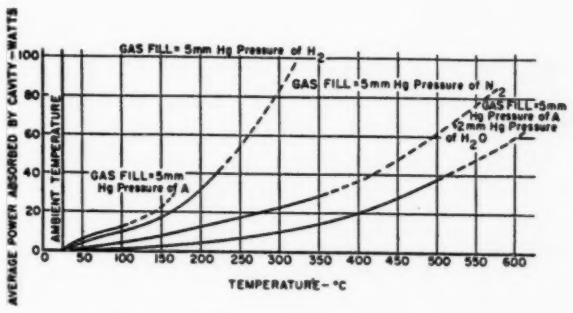

AVERAGE POWER ABSORBED BY CAVITY-WATTS
AMBIENT TEMPERATURE
GAS FILL=5mm Hg Pressure of H_2
GAS FILL=5mm Hg Pressure of N_2
GAS FILL=5mm Hg Pressure of A
GAS FILL=5mm Hg Pressure of A +2mm Hg Pressure of H_2O
TEMPERATURE-°C

FIG. 5. Average power absorbed by the cavity *versus* gas temperature as a function of gas fill for cavity with window $Q=0.6$. Value of window Q is for the doubly-loaded condition. The dashed portion of the curves are based upon extrapolated data.

the gas temperatures which the windows can successfully withstand. Type-707 glass is generally used since it has the lowest dielectric loss of any of the common glasses, and its coefficient of expansion is reasonably well matched to that of Kovar over a wide temperature range. The strain point, which is defined as the highest temperature from which strain-free glass can be cooled quickly without introducing permanent strain, has a value of 455 deg. C for type-707 glass. It can be seen from Fig. 3 that this temperature is exceeded for the argon and water vapor gas fill before the 2-megawatt peak power level is reached and that it does in fact result from a peak power level of somewhat less than 0.7 megawatt for the cavity with a window Q of 0.6. It is significant that ATR tubes which use an argon and water vapor gas fill and this low Q window are comparatively short-lived if the system peak power level exceeds 0.7 megawatt. In addition, the glass is subjected to severe mechanical strains because of the temperature differential of several hundred degrees between the inside and outside surfaces of the window, which is 0.060 in. thick. Consequently, a temperature gradient exists which sets up tensile stresses of several thousand pounds per sq in. because the hot inner

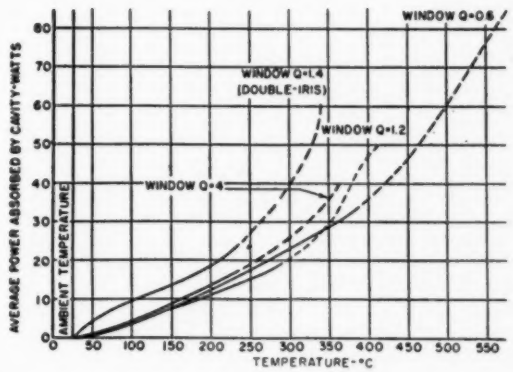


FIG. 6. Average power absorbed by the cavity *versus* gas temperature as a function of window Q for nitrogen-gas fill. Values of window Q are for the doubly-loaded condition. The dashed portion of the curve is based upon extrapolated data.

surface expands and pulls the cooler outer surface into tension.[4]

Efforts to correlate the relationship between the effect of the gas fill on the resultant temperature attained and power absorbed by cavity, and gas properties such as specific heat, viscosity, heat conductivity, and ionization potential were unsuccessful. However, an analysis was made to determine the mathematical expression to which the curves shown in Fig. 3 could best be fitted. It was found that expressions of the form $y=ax^n+b$ would fit these curves to within 6 percent in all cases. The data shown in Fig. 5 does present a definite qualitative indication as to the ability of the gas fills investigated to dissipate the heat generated by the rf discharge.

The power absorbed by any one cavity at the high power levels increases linearly with line power. An indication of this can be seen from Table I. This conforms to the well-known characteristic behavior of this type of tube wherein the db arc loss is primarily a function of the window Q and remains very nearly constant for large variations in line power at high power levels.

The curves shown in Fig. 6 all bunch together at the lower temperature levels, which correspond to the lower power levels. This seems to be an indication that the window itself has little effect upon the ability of the cavity to dissipate heat at lower power levels, but is a contributing factor at higher power levels. The cavities with the higher Q single-iris windows are capable of dissipating the most heat at the higher power levels; a result which would be anticipated since the metal conducting walls of the cavity are in closer proximity to the discharge.

The apparent inconsistencies in these data such as the intersecting of the curves, may result from the fact that it was necessary to use extrapolated data in arriving at the results presented in these curves. Slight deviations in this extrapolated data will cause the inconsistencies noted.

It should be borne in mind that the discussion of the heat-dissipating characteristics of the gas and of the various cavities is based upon a condition wherein the power absorbed by the cavity is under consideration, and that equal amounts of power absorbed by different cavities do not necessarily correspond to the same incident line power.

[4] E. B. Steinberg, Elec. Mfg. 44, No. 5, 82–88 (1949).

CONCLUSIONS

The investigation discussed in this paper has shown that the method used in measuring the gas temperature resulting from an rf discharge in an ATR type of cavity is valid. This temperature may be sufficiently high so that the glass in the resonant window of the cavity is incapable of withstanding the heat generated by the discharge. The electrical characteristics of the cavity due to the window geometry and gas-fill properties, and the ability of the glass window to withstand the mechanical strains induced by the temperature of the rf discharge are of considerable importance in determining the maximum line power levels at which this type of cavity can be successfully operated. A double-iris type of window, by combining good mechanical strength with a geometry that results in a reasonably low Q cavity, can operate at higher power levels and dissipate more heat than comparable cavities with single-iris windows.

Qualitative results indicate that cavities with higher Q, single-iris windows are more effective than cavities with lower Q, single-iris windows in dissipating heat at high line power levels; but that the window has comparatively little effect at low power levels.

Experimental evidence is presented which shows that an argon-gas fill has the greatest "effective heat conductivity," inasmuch as it was the most effective in dissipating the heat generated by the rf discharge of any of the gas fills investigated. Hydrogen, nitrogen, and argon and water vapor follow in that order.

ACKNOWLEDGMENT

The authors wish to thank Mr. E. V. Edwards, who supervised the assembly and processing of the cavities tested.

Letters to the Editor

On the Diffraction of X-Rays by a Twisted Crystal Lattice

S. S. Hsu
Department of Metallurgy, University of Notre Dame, Notre Dame, Indiana
(Received December 2, 1953)

GUINIER and Tennevin[1] have devised a transmission Laue method by which very small amounts of lattice bending can be measured. Cullity and Julien,[2] by placing a Soller slit in the incident beam, extended the Guinier-Tennevin method in such a way that they were able to measure the torsion of a crystal lattice.

The use of a Soller slit, with its plates horizontal, in the beam issuing from a vertical line source of x-rays was shown to produce a set of virtual sources, one at the center of each cell of the Soller slit.[2] These sources lie at the intersections within the Soller slit of the rays shown in Fig. 1 (a). (In this schematic drawing, the incident rays are in the plane of the drawing, the crystal has the form of a thin plate with its reflecting lattice planes almost parallel to the plane of the drawing and the diffracted rays lie somewhat out of this plane.) Radiation from such a set of sources is diffracted by a crystal twisted about a vertical axis to form a Laue spot, on the equator line of the film, which has the appearance of I_p in Fig. 1 (a). Here I_p is the front view of the reflection registered on the film F. It consists of a set of parallel lines or striations inclined at an angle γ to the vertical, and $\tan\gamma$ is directly proportional to the degree of torsion of the crystal.

Guinier[3] suggested that quite similar results might be obtained by using a grid of fine horizontal wires in place of the Soller slit. This suggestion was tested with a grid of 20 steel wires, each $\frac{1}{4}$ mm in diameter and so spaced that there was an opening of $\frac{1}{4}$ mm between adjacent wires. Since the wire grid has no ability to limit the beam, a lead sheet with an aperture 2 mm high and 0.5 mm wide was placed immediately in front of the specimen to limit the size of the diffraction spots and prevent blackening of the film by the incident beam. Figure 1 (b) shows the experimental arrangement: the source-crystal and crystal-film distances were of the order of 15–20 cm. Figure 2 is typical of the Laue spots observed under these conditions; it was obtained from a rectangular quartz crystal plate twisted about its longitudinal axis by 0.1 degree/mm. The appearance of this reflection is similar to that obtained with a Soller slit. It contains striations and these are inclined to the vertical at precisely the angle computed from the known torsion of the crystal.

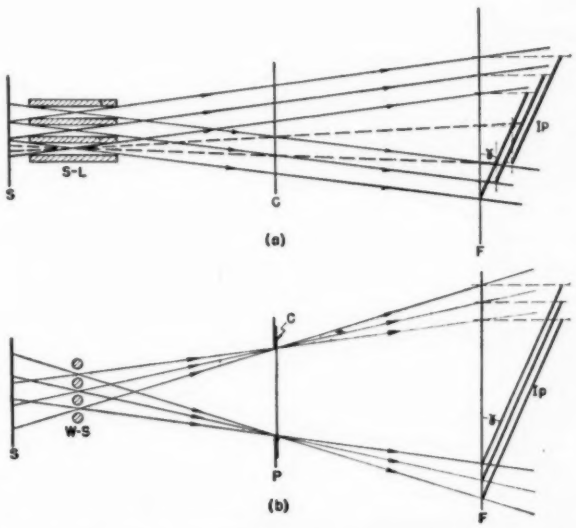


FIG. 1. Diffraction geometry for incident beam passing through (a) a Soller slit and (b) a set of equally spaced wires. S = x-ray source, S–L = Soller slit, W–S = wire grid, P = aperture, C = crystal, F = film, I_p = projection of image.

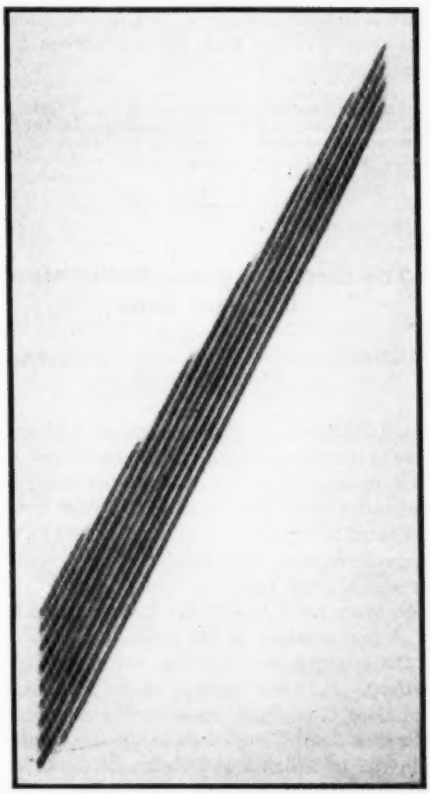


FIG. 2. Transmission Laue spot from twisted quartz crystal. 3×.

There is a difference, however, between the striation arrangement obtained with a grid [indicated schematically by I_p in Fig. 1 (b)] and that obtained with a Soller slit. The ends of the striations produced by a grid lie on two vertical lines on the film while those produced by a Soller slit are shifted horizontally as well as vertically with respect to one another. This difference in striation arrangement is due essentially to differences in the way the crystal is illuminated by virtual sources in each case. Consider the lowermost rays from each source. When a Soller slit is used, the lowermost rays from each virtual source are necessarily parallel: they, therefore, strike the crystal at successively higher points, that is, at successively larger Bragg angles since the crystal is twisted about a vertical axis. These rays are, therefore, diffracted to points on the film which are displaced, relative to one another, not only vertically but also horizontally. On the other hand, when a grid is used there are also virtual sources in the openings between adjacent wires, but the divergence from each source is limited only by the aperture in front of the crystal. The lowermost rays from each source, therefore, strike the same point on the crystal as shown in Fig. 1 (b), and therefore at the same Bragg angle. Accordingly, they are diffracted to points on the film which are separated only vertically and not horizontally.

It may be concluded that a wire grid is just as effective as a Soller slit, when used in conjunction with a line source of x-rays, in detecting and measuring torsional strain in a crystal lattice, although the appearance of the Laue spots obtained by the two methods differ in detail.

The writer is indebted to the U. S. Office of Naval Research for financial support of this work and to Professor B. D. Cullity for helpful suggestions.

[1] A. Guinier and J. Tennevin, Acta. Cryst. 2, 133 (1949).
[2] B. D. Cullity and Carl A. Julien, J. Appl. Phys. 24, 541 (1953).
[3] A. Guinier, private communication.

The Electromagnetic Field near a Dielectric Lens

D. C. Hogg
*The Eaton Electronics Research Laboratory, McGill University, Montreal, Canada**
(Received November 23, 1953)

MEASUREMENTS at a wavelength of 1.25 cm have been made in the near field of a doubly convex polystyrene lens. The lens was mounted in a circular aperture thirty-three wavelengths in diameter which was cut in a thin, plane metallic screen; it was illuminated by a point source placed at its focal point.

All measurements were made parallel to the screen and on the side of the screen away from the source. Figure 1 shows three sets of these taken at distances $Z_0=3.5\lambda$, 4λ, and 5λ from the screen. $|E_y|^2$ is a measure of the y component of the electric field along the magnetic diameter (the incident field polarized in the y direction). $|E_x|^2$ is a measure of the x component of the electric field along the electric diameter (the incident field polarized in the x direction). Comparison of the three sets shows that the fields in the two planes of polarization differ considerably near the lens surface, whereas they assume the same form fairly rapidly with increasing Z_0.

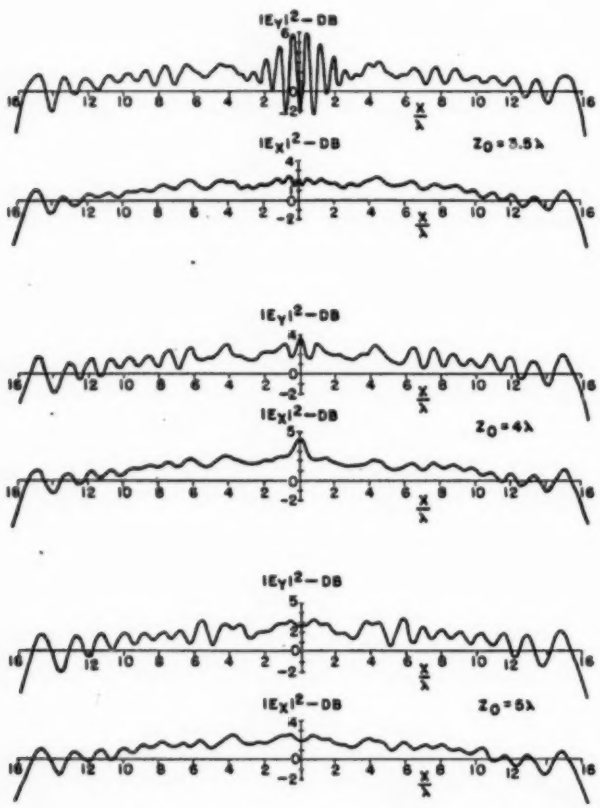


FIG. 1. The near field of a polystyrene lens.

The first pattern in the figure shows that very closely spaced field intensity variations exist along the magnetic diameter near the optic axis. At this distance ($Z_0=3.5\lambda$), the measuring probe just brushes the convex surface of the lens near the optic axis. The separation of the successive maxima is found to correspond closely to the wavelength in the dielectric material. These measurements are unlike those obtained when no dielectric is present and when the field near a circular aperture alone is examined,[1,2] in which case the spacing between successive maxima is of the order of one free-space wavelength. Although no quantitative explanation can be offered for this field, a possible interpretation is that the virtual sources along the aperture edge which were suggested by Andrews,[1] in this case give rise to wavelets within the dielectric. Under the assumption of radiation from the edge of the aperture, calculation from the geometry of the lens shows that close to the optic axis, this secondary radiation is totally reflected; this radiation in turn gives rise to a surface wave just outside the dielectric which interacts with the radiation passing directly through the lens. In addition to the fact that this type of field should have a spatial period equal to the wavelength within the dielectric, it should die out at about one wavelength from the dielectric surface as has been observed.

A similar rapidly fluctuating field was measured near a thin polystyrene disk mounted in a circular aperture when the aperture was irradiated by a plane wave.

This research was done under Grant DRB 176 made by the Defence Research Board of Canada to the Eaton Electronics Research Laboratory, McGill University, Montreal, Canada.

* Now of the Bell Telephone Laboratories, Holmdel, New Jersey.
[1] C. L. Andrews, J. Appl. Phys. 21, 761 (1950).
[2] D. C. Hogg, J. Appl. Phys. 24, 110 (1953).

Application of Perturbation Theory to the Traveling Wave Tubes

M. W. Muller and W. L. Beaver
Varian Associates, Palo Alto, California
(Received November 23, 1953)

IN developing the theory of the TWT (traveling wave tube) it is a common procedure to derive a differential equation governing the behavior of some convenient variable such as rf current or electric field. With the assumption of plane-wave solutions this differential equation is reduced to an algebraic (secular) equation whose roots are the several propagation constants. The degree of the algebraic equation is determined by the nature of the problem and by the simplifying assumptions which are made in deriving the differential equation. Thus, if the problem is considered one-dimensional, the equation is a quartic; if transverse electron motion is allowed (as, for instance, in the magnetron amplifier) it is of the fifth degree; if the circuit is nonpropagating (as in an easitron or in a drift tube) it is a quadratic; and so forth.

In the absence of any coupling between the electrons and the circuit, the roots of the secular equation are easily obtained and are very simple in form. In this case the solutions describe unperturbed propagation of waves along the circuit, and unperturbed plasma waves in the electron beam. When interaction between beam and circuit is introduced, the secular equation may become somewhat awkward and the roots are generally obtained by an approximation based on the smallness of the interaction. This procedure is usually characterized by a "gain parameter" which is assumed to be small.

It is the purpose of this note to show that this procedure can be described by a simple and general formulation which correlates a number of previously known results.

The secular equation may be written

$$f(u_i,\epsilon)=0$$

where u_i is the ith root (propagation constant) and ϵ is the interaction term. The u_i are assumed known if $\epsilon=0$, and are denoted u_{i0}.

Now by Taylor's theorem

$$f(u_i,\epsilon)=f(u_{i0},0)+\frac{\partial f}{\partial u_i}(u_i-u_{i0})+\frac{\partial f}{\partial \epsilon}\epsilon+\frac{1}{2}\frac{\partial^2 f}{\partial u_i^2}(u_i-u_{i0})^2+\frac{\partial^2 f}{\partial u_i \partial \epsilon}\epsilon(u_i-u_{i0})+\frac{1}{2}\frac{\partial^2 f}{\partial \epsilon^2}\epsilon^2+\cdots,$$

and if we denote the incremental propagation constants by $\xi_i=u_i-u_{i0}$, we have an equation for the ξ_i,

$$\frac{\partial f}{\partial u_i}\xi_i+\frac{\partial f}{\partial \epsilon}\epsilon+\frac{1}{2}\frac{\partial^2 f}{\partial u_i^2}\xi_i^2+\frac{\partial^2 f}{\partial u_i \partial \epsilon}\xi_i\epsilon+\frac{1}{2}\frac{\partial^2 f}{\partial \epsilon^2}\epsilon^2+\cdots=0.$$

Since ϵ is small, and from physical continuity we may assume that ξ is small, we can neglect all but the leading terms and we obtain

$$\xi_i=-\left(\frac{\partial f}{\partial \epsilon}\Big/\frac{\partial f}{\partial u_i}\right)\epsilon,$$

unless u_{i0} is a multiple root. If u_{i0} is an n-fold root of the secular equation, the lowest nonvanishing partial derivative with respect to u_i is $\partial^n f/\partial u_i^n$ and the incremental propagation constant is given approximately* by

$$\xi_i^n=-\frac{\dfrac{\partial f}{\partial \epsilon}}{\dfrac{1}{n!}\dfrac{\partial^n f}{\partial u_i^n}}\epsilon.$$

Now in a conventional TWT operated at synchronism there is a triple root corresponding to three degenerate forward waves (two space charge waves and a circuit wave), and hence the incremental propagation constants are proportional to $\epsilon^{\frac{1}{3}}$ (and hence to $I_0^{\frac{1}{3}}$ where I_0 is the beam current), while the incremental propagation constant for the backward wave is proportional to ϵ. Similarly in the case of the growing waves of a magnetron amplifier or an easitron or a resistance wall amplifier in which the degeneracy is twofold, the incremental propagation constant is proportional to $\epsilon^{\frac{1}{2}}$ etc.

* If two or more roots are nearly degenerate, as for instance in the theory of the TWT including space-charge forces, the coefficient of the lowest-order term in ξ will be small and a better approximation can be obtained by including higher-order terms.

Regrowth of Silicon through a Low Melting Zone of Silicon-Gold Eutectic*

J. N. Carman, P. E. Stello, and C. A. Bittmann
Hughes Research and Development Laboratories, Culver City, California
(Received November 23, 1953)

EXPERIMENTS have been reported on silicon crystals grown from molten alloys of silicon dissolved in aluminum, silver, and zinc by von Wartenberg,[1] in gallium and indium by Keck and Broder,[2] and in tin by Gross.[3] The present note deals with regrowth of silicon through a low-melting zone of silicon-gold eutectic.

Zone melting techniques described by Pfann[4] were employed. By using a seed of silicon, followed by a zone of silicon-gold eutectic,† followed by a source of silicon, regrowth of silicon on the seed occurs. Silicon crystals, substantially free of occluded gold, have been produced up to ¾ in. in length. This material was "P" type of 4 ohm-cm resistivity and a minority carrier lifetime of 2 μsec. The best results were obtained when the rate of movement of the heater was 0.001 in./hr.

The process described provides a method for producing silicon and an experimental method for studying the process of regrowth used in the production of fusion-type junctions.

* This work, supported in part by U. S. Signal Corps Contract No. DA 36-039 SC 42574, was described by C. A. Bittmann at the National Conference on Tube Techniques sponsored by the Research Development Board Subpanel on Tube Techniques, New York, October 15, 1953.

† The eutectic of the silicon-gold system occurs at a temperature of 370°C and a composition of 6 percent silicon, 94 percent gold [M. Hansen, *Der Aufbau der Zweistofflegiernungen* (Julius Springer, Berlin, 1936)].

[1] H. v. Wartenberg, Z. anorg. u. allgem. Chem. **265**, 186 (1951).
[2] P. H. Keck and J. Broder, Phys. Rev. **90**, 521 (1953).
[3] A. J. Gross, J. Metals (September, 1953).
[4] W. G. Pfann, Trans. Am. Inst. Mining Met. Engrs. **194**, 747 (1952); J. Metals (February, 1951).

Erratum: Approximate Solutions of the Space-Charge Problem for Some Unusual Electrode Geometries

Henry F. Ivey
Westinghouse Electric Corporation, Bloomfield, New Jersey
[J. Appl. Phys. **24**, 1466 (1953)]

AN error in the last paragraph on page 1471 of the above paper has been pointed out by Dr. Alan Hazeltine. This paragraph should be replaced by the following:

It is of interest to note that not all problems involving spherical emitters may be solved by the technique discussed here. As an example may be mentioned the case of two spherical electrodes, of radii r and R, neither of which is inside the other. As the distance between the spheres is increased, the capacitance between them approaches as a limit $1/C=(1/r)+(1/R)$, so that C is always less than r. This situation cannot be represented by two concentric spheres of radii r and R' and with the inner sphere the emitter $(R'>r)$, because the capacitance of this system is given by $1/C'=(1/r)-(1/R')$, so that C' is always greater than r.

The additional comment may be made concerning this paper that many of the electrode arrangements considered depart so markedly from the equivalent symmetrical concentric arrangement that the validity of the Matricon and Trouvé premise, i.e., that the shape of the equipotential surfaces in the immediate neighborhood of the emitter is the same in the two cases, may be questioned. However, experimental data to indicate the accuracy of the approximation involved are not at hand.

Errata: The Vector Wave Function Solution of the Diffraction of Electromagnetic Waves by Circular Disks and Apertures. II. The Diffraction Problems

[J. Appl. Phys. **24**, 1224–1231 (1953)]

Carson Flammer
Stanford Research Institute, Stanford, California

I WOULD like to call attention to the following typographical errors in my paper:

In Eq. (11), $R_{1n}^{(3)}(-ic,i\xi)$ should be replaced by $R_{1n}^{(3)\prime}(-ic,i\xi)$. In Eq. (13), $R_{0n}^{(1)\prime}(-ic,o)$ should be replaced by $S_{0n}^{(1)\prime}(-ic,o)$.

Errata: Sudden Change in the Arrangement of Crystallites in Thin Evaporated Bismuth Films

[J. Appl. Phys. *25*, 125 (1954)]

Lorn L. Howard
Conducting Thin Films Laboratory, Department of Electrical Engineering, University of Illinois, Urbana, Illinois

LINE 11 should read: "distance of 30 cm. This bismuth film was of the order of 1000A. . . ."

Page 127, line 2 should read: "film of approximately the same thickness as shown by. . . ."

Announcement

International Symposium

Information Networks is the topic of the third of a series of annual international symposia to be held April 12, 13, and 14, 1954, at the Engineering Societies Building, 33 West 39th Street, New York 18, New York.

The symposium will deal with network theory, particularly network synthesis as it is influenced by the newer concepts developed in information and general communication theory. In particular, the first part will concentrate upon the performance of networks and their design for specific types of information such as pulses, pulse modulation, statistical inputs, etc. The second part will concentrate upon the generalized network concepts and their application to computer and switching systems, neuron networks, and optical systems. American and European authorities, who have made original contributions to the art, will participate.

The cooperation of the Professional Group on Circuit Theory of the Institute of Radio Engineers and the co-sponsorship of the U. S. Office of Naval Research, the Air Force Office of Scientific Research, and the Signal Corps permits this symposium to be held without admission charge or registration fee. "Proceedings of the Symposium on Information Networks" will be published by October, 1954, at a cost of four dollars ($4.00) per copy. A cloth-bound edition will also be available at additional cost. Members of the Professional Group on Circuit Theory of the Institute of Radio Engineers may obtain a copy at a saving of one dollar ($1.00). Orders for the "Proceedings," accompanied by check or money order, made out to Treasurer, Network Symposium, will be accepted in advance. Copies of the detailed program, hotel accommodation information, and registration forms are available on request. All correspondence should be addressed to:

Polytechnic Institute of Brooklyn
Microwave Research Institute
55 Johnson Street
Brooklyn 1, New York

All persons interested in attending are urged to register early.